PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 17 August 15, 1931 Number 8

THE PHYSICAL BASIS OF MODERN HYDROGRAPHIC SURVEYING

By A. L. Shalowitz

Cartographic Engineer, U. S. Coast and Geodetic Survey

Read before the Academy April 28, 1931

Hydrographic surveying, as it is now practised in the Coast and Geodetic Survey, is so radically different from what it was a few years ago that to fully grasp the significance of the new order, one must be acquainted with the limitations and shortcomings of the old methods. To say that it is now possible to conduct major surveying operations in areas hitherto excluded because of the prohibitive cost involved, is only giving one side of the picture. It is now actually possible to survey such areas with an accuracy of detail sufficient to meet the needs not only of the navigator but of the scientific investigator as well. This remarkable advance, during the post-war period has come about principally through the application of the physical sciences, particularly in the field of acoustics, to the problems of the hydrographic engineer.

The possibility of utilizing sound as a means of measuring ocean depths and distances, was recognized long before this period, but no *practical method* had been evolved for meeting the exacting demands of modern hydrographic surveys. The scientific investigations made during the war were quickly focused on peace-time needs by the leading maritime nations of the world, resulting in this country in the development of the sonic depth finder by the United States Navy and the fathometer by the Submarine Signal Corporation.

While other countries have developed other types of echo-sounding machines, the underlying principle of all is the same, involving the measurement of the elapsed time between the emission of a sound impulse from the vessel and the return of the echo from the bottom. One-half this interval multiplied by the velocity of sound in sea water determines the depth under the surveying vessel. In the fathometer, the revolutions of a disc driven by a constant speed motor measures the elapsed time between the outgoing and incoming signals. A graduated dial calibrated for a standard velocity translates this elapsed time into depth units. But since

velocity varies with the temperature, salinity and pressure of the water, these variables must first be determined before an accurate sounding can be obtained.

In the Coast and Geodetic Survey temperatures and salinities are measured at well-distributed points over the area to be surveyed, and the mean velocities within any range of depths are ascertained from theoretical velocity tables, to within very narrow limits.

The advantages of this method of sounding over former methods, particularly in deep water, are too well established to be heralded at this late date. Suffice it to say that not only is it possible to obtain a continuous profile of the ocean floor, but a vessel can now, while steaming along at

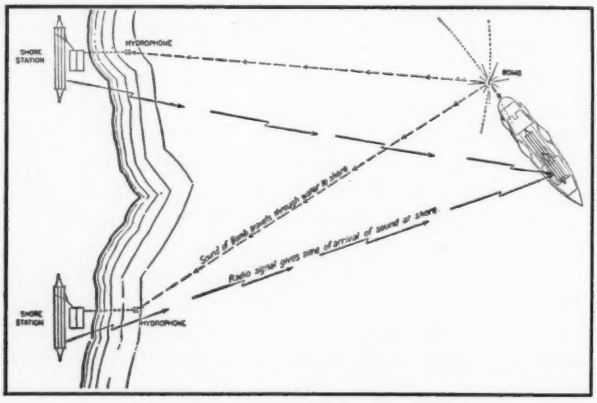


FIGURE 1
Radio acoustic position finding.

full speed, obtain a sounding of 2000 fathoms in five seconds, which formerly took from 40 to 50 minutes.

Determination of depth, however, is only one element in the prosecution of a hydrographic survey. For unless we also know the positions of these depths on the surface of the water, that is, their relation to fixed points on shore, a properly coördinated survey is impossible, and the results would be of little value as an aid to navigation or as a basis for scientific study.

Distances from shore can be determined by various means, the simplest of which is the measuring of angles between three known objects on shore or three well-located buoys offshore. This method, of course, depends for its usefulness upon the condition of the weather and upon the limitations imposed by the earth's curvature.

Beyond this limit recourse must be had to other expedients, such as "dead reckoning" and astronomic observations. Both of these methods,

however, are also subject to very definite limitations; as for example, the effect of current and wind on the course and distance, the errors of the compass, the state of the weather, the condition of the sea horizon; all of which make accurate, related surveys impossible, and in areas of considerable relief often results in an improper representation of submarine features.

It was these recognized difficulties, together with the natural evolution of hydrographic methods, that prompted the Coast and Geodetic Survey to undertake the development of an acoustic method for locating the position of a survey ship; and in 1923, in collaboration with the Sound Laboratory of the Bureau of Standards, to develop what is known today as the "Radio Acoustic Method of Position Finding."

Briefly stated, the method consists in exploding a small charge of T.N.T. near the surveying vessel, the time of explosion being recorded on a chronograph aboard the vessel. The sound wave being spherical is propagated in all directions from the center of disturbance and after the lapse of a certain interval reaches the hydrophones located at predetermined stations near shore or at suitable points offshore. Here it is transformed into an electrical impulse, which operates an automatic key at the shore station and a radio signal is sent back to the ship where a second mark is made on the same chronograph. Since all the electrical operations are automatic and instantaneous, the distance between the two marks represents the time taken for the sound wave to travel from the vessel to the hydrophone. If then the horizontal velocity of sound through the water is known, a measure of the distance to the vessel is afforded.

Because such methods can be used by day or night and under all weather conditions, it is of particular importance economically. And when used in conjunction with echo soundings, greater accuracy in the work is not only assured, but an area can now be surveyed with a degree of detail commensurate with the needs of the area, and yet at a cost far below that of former methods. To illustrate: it took a vessel 87 hours in 1922 to survey an area that in 1929 took but 44 hours, and yet five times the number of soundings were taken.

Is it surprising then that during the last few years such rapid strides have been made in completing much needed surveys along the Pacific coast?

And one needs but to examine a modern survey of a deep water area to realize what these new methods have meant in the way of delineation of important underwater configurations.

A striking example of the value of such methods was brought home to the Bureau last summer when the important but difficult survey of Georges Bank was undertaken. This bank is of vital concern to a vast fishing industry and of outstanding importance to transatlantic shipping, lying close to the principal westbound steamer lane between Europe and the

United States. It covers an area twice the size of the State of Massachusetts and its eastern edge is almost 200 miles from the nearest point of land on the New England coast. Because frequent fogs and thick weather in this locality necessitate a vessel's recourse to soundings for fixing its posi-

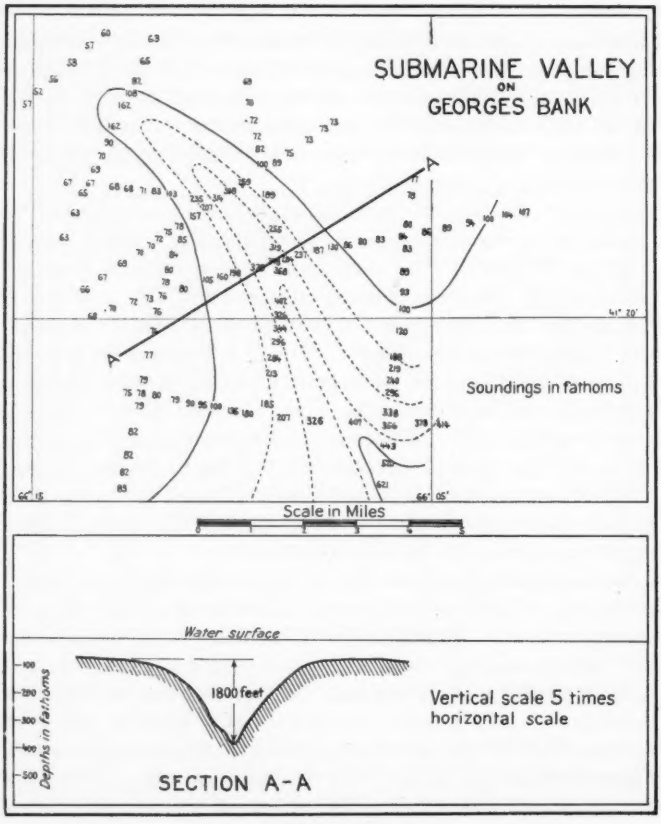


FIGURE 2

Submarine valley on Georges Bank (from unadjusted field sheet).

tion, the need for a detailed, coördinated and reliable survey of the area had long been recognized.

The acoustic methods of surveying have made such results attainable and the undertaking practicable. Already the survey of this bank has developed the existence of an uncharted submarine valley, two miles wide, eight miles long and 1800 feet deep lying directly in the transatlantic

steamer lane and ideally oriented for vessels reshaping their course for Nantucket Shoals Light Vessel. If this submarine feature should prove unique, its value as an aid to navigation cannot be overestimated, and standing alone will have more than justified the entire cost of the project.

It is expected that when all the work on this bank is completed, a wealth of information will be unfolded, of interest alike to the geologist, physiographer, oceanographer, hydrographer and mariner.

But while the acoustic method of surveying has overcome many difficulties and increased our accomplishments, it has also brought many new problems. While most of these have been solved, sometimes by adopting practical expedients, as in the case of slope corrections for echo soundings (see Special Publication 165, Coast and Geodetic Survey) many more remain to be dealt with. Of considerable importance at the present time is the determination of the path followed by the sound wave in horizontal transmission in sea water.

It will be recalled that the velocity of sound in sea water varies with the temperature, salinity and pressure of the water. When dealing with echo soundings, where the sound penetrates all the layers from surface to bottom, the problem resolves itself into determining the mean velocity for the entire column of water through which the sound wave passes. Theoretical velocity tables have been computed based on Newton's fundamental equation,

$$\text{Velocity} = \sqrt{\frac{\text{Elasticity of the Medium}}{\text{Density of the Medium}}}$$

giving the velocity for various conditions of temperature, salinity and pressure and if the existing conditions are known the actual velocity of the sound can readily be determined.

In Radio Acoustic Position work, however, the problem of determining the mean velocity of the sound wave between vessel and shore is quite different. Here we are concerned not with vertical transmission but with horizontal transmission. And because the velocity varies between surface and bottom, unless we know the path followed by the sound wave, we cannot determine the velocity to be used for computing the position of the vessel.

Figure 3 represents a cross-section of an area off the Oregon coast from the shore to the 50-fathom depth curve. The variations in temperature between surface and bottom for this section are shown in the insert. The velocities indicated are mean theoretical velocities, computed from the temperature, salinity and pressure data for the three assumed paths of the sound wave from bomb to hydrophone. It will be seen that the difference between the surface velocity and the bottom velocity is about 22 meters

per second which at a distance of 60 nautical miles might introduce an error in position of approximately 1600 meters if we made a wrong assumption as to the path of the wave. So it is quite evident that the problem is of more than mere academic concern.

It is of course possible to determine a velocity experimentally, when in close proximity to shore signals, by locating the position of the vessel by the ordinary methods of observing angles and simultaneously measuring the time for the sound impulse to reach the hydrophone. By scaling the distance on the survey sheet from the vessel to the hydrophone and dividing by the time interval we obtain the mean velocity of the sound wave be-

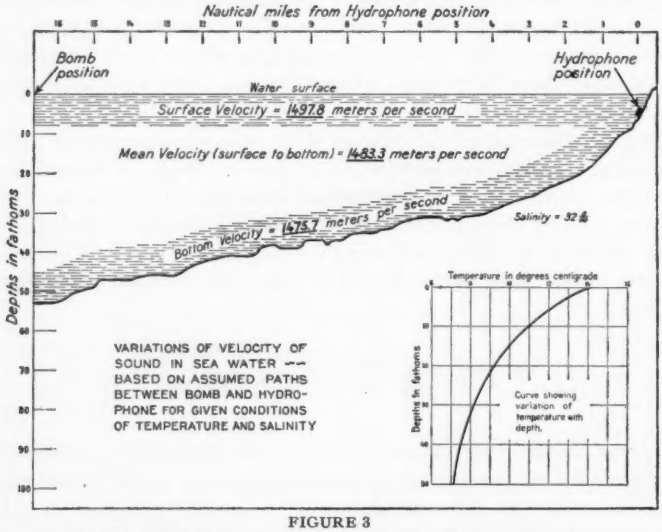

Nautical miles from Hydrophone position
Bomb position
Hydrophone position
Water surface
Surface Velocity = 1497.8 meters per second
Mean Velocity (surface to bottom) = 1483.3 meters per second
Bottom Velocity = 1475.7 meters per second
Salinity = 32‰
Depths in fathoms
VARIATIONS OF VELOCITY OF SOUND IN SEA WATER — BASED ON ASSUMED PATHS BETWEEN BOMB AND HYDROPHONE FOR GIVEN CONDITIONS OF TEMPERATURE AND SALINITY
Temperature in degrees centigrade
Depths in fathoms
Curve showing variation of temperature with depth.

FIGURE 3

Variation of horizontal velocity of sound with assumed paths.

tween bomb and hydrophone. This velocity can then be used when working in a locality where the physical conditions of the water are approximately the same. But when the field of work is far removed from shore control some other means must be found for determining the most probable velocity of the area. This was the problem with which we were confronted when the work on Georges Bank was undertaken. In view of the vast amount of data that had been accumulated by the Coast and Geodetic Survey in its sound ranging work on the West Coast, it seemed desirable to approach the problem from the observational rather than the theoretical point of view.

A detailed study was, therefore, made of all the experimental velocity data along the Pacific coast and in Alaska, with a view to determining the rela-

tion existing between experimental values and theoretical values based on assumed paths. As the sound wave could pass through an infinite number of physical conditions between surface and bottom in its journey from bomb to hydrophone, it was decided at the outset to limit the investigation to three principal considerations:

1. The relationship of experimental velocities to theoretical *surface* velocities.

2. The relationship of experimental velocities to *mean* theoretical velocities between surface and bottom.

3. The relationship of experimental velocities to theoretical *bottom* velocities.

It should be emphasized at this point that while velocities have been

Group 33–37 J' (H. 4636)—Coast of Oregon—September 24, 1926.

To KGAM—Test in 88 fathoms—Measured time 28 seconds

Sound passes through 74 fathoms for 0.7 distance
Sound passes through 45 fathoms for 0.15 distance
Sound passes through 32 fathoms for 0.10 distance
Sound passes through 18 fathoms for 0.05 distance

Temp. at surface = 13.1°
Temp. at 18 fms. = 8.6
Temp. at 32 fms. = 7.2
Temp. at 45 fms. = 7.0
Temp. at 74 fms. = 6.8
} From serial temps. taken September 29, 1926, about five miles to westward

Salinity = 33.4

Velocity at surface = 1495.9 meters per sec.

Bottom Velocity			*Mean Velocity* (Surface to bottom)		
Velocity at 8.6°	= 1480.3	Corrected for	Mean temp. at 18 fms.	= 11.1°	From
Velocity at 7.2	= 1475.4	salinity and	Mean temp. at 32 fms.	= 9.6	serial
Velocity at 7.0	= 1475.9	for pressure	Mean temp. at 45 fms.	= 9.0	temps
Velocity at 6.8	= 1475.0	at the respec-	Mean temp. at 74 fms.	= 8.2	as
		tive depths			above.
Mean velocity (weighted)	= 1475.4		Mean temp. bomb to hyd.	= 8.6°	
			Mean depth bomb to hyd.	= 31 fms.	
			Mean velocity	= 1480.7	

Experimental Velocity = 1475.4 (mean of 5 good obs.)

FIGURE 4
Typical computation of theoretical velocities

measured in the past by various investigators with a high degree of accuracy and isolated comparisons made with theoretical values, never has so much data covering such a variety of conditions been available for investigative purposes.

Actually the present study covered the experiences of eight survey parties over a period of four years, and included experiments in areas extending from the cold waters of the Gulf of Alaska to the warmer waters of northern California. The comparisons cover tests made in depths vary-

ing from 20 to 250 fathoms and at distances ranging from three to 60 nautical miles from the hydrophone. The cases selected for comparison are random ones and chosen without regard to probable results. The deductions to be drawn are therefore free from any bias of specially selected observations.

For each experimental velocity a computation similar to the one shown in figure 4 was made. The theoretical velocities were determined in the following manner: The vertical section between vessel and hydrophone was first subdivided into zones of various depths, the number depending upon the regularity or irregularity of the bottom. Velocities were then computed for each zone from the temperature, salinity and pressure data for that zone. The mean velocity was then obtained by weighting these various velocities in accordance with the ratio that each zone bore to the entire distance. In computing the theoretical velocities, the British Admiralty tables (H.O. 282) were used exclusively, being particularly well adapted for sound ranging work. Corrections on account of the variation of the force of gravity with latitude were entirely disregarded, since the corrections for the depths involved would not exceed one-tenth meter per second.

Figure 5 gives comparisons of experimental velocities with theoretical velocities for various paths. In studying this table it should be borne in mind that although 51 comparisons only are listed, in reality many more are represented, for each experimental value may be the value of a single determination or it may be the mean of a group of determinations, frequently as many as five. Wherever it was possible to use a number of observations as a basis, as in cases where the tests were made close together and the sound passed through approximately the same depths under the same conditions of temperature and salinity, this was done, since such procedure eliminated small accidental errors in the measured velocities.

The temperature data on which the theoretical velocities were based were selected from observations taken close to the day on which the velocity tests were made and in approximately the same locality. In a number of cases the mean season curve for the various depths had to be resorted to. Since temperature is the predominating factor, at least in shoal water, in the determination of sound velocity in sea water, theoretical velocities computed from such data must of necessity be somewhat doubtful. It is nevertheless well to remember that any temperature variation from the mean will manifest itself more in the surface and mean velocities rather than in the bottom velocities, since the latter depend on depths that are usually below the region of high temperature fluctuations.

Figure 5 shows that in all the cases considered the theoretical surface velocities exceed the measured velocities by an average of 12.3 meters per second, that the mean velocities (between surface and bottom) exceed the

measured velocities by an average of 3.2 meters per second, *but that the bottom velocities average only one meter per second less than the corresponding measured velocities.* It should be noted in this connection that the comparisons in Shelikof Strait are the only ones where the average difference for the group between the experimental velocity and the theoretical mean velocity is less than the average difference for the bottom velocity. No explanation for this exception could be adduced from the available data. It is, however, significant that this group is the only one that indicates the effective sound energy as passing through temperature conditions higher than the mean. In all other cases the reverse holds true. It would, therefore, appear reasonable to reject these comparisons as erratic. And the

EXPERIMENTAL VELOCITY (E)	SURFACE VELOCITY (S)	MEAN VELOCITY (M)	BOTTOM VELOCITY (B)	E − S	E − M	E − B
		Gulf of Alaska				
1469.8	1490.9	1478.1	1469.2	−21.1	−8.3	+0.6
1467.4	1483.0	1477.2	1470.2	−15.6	−9.8	−2.8
1471.7	1483.0	1480.1	1473.2	−11.3	−8.4	−1.5
1471.0	1483.0	1474.8	1469.2	−12.0	−3.8	+1.8
1470.0	1484.8	1477.0	1469.5	−14.8	−7.0	+0.5
1468.9	1495.1	1476.9	1468.8	−26.2	−8.0	+0.1
1469.2	1483.7	1474.6	1469.3	−14.5	−5.4	−0.1
				Av. −16.5	Av. −7.2	Av. −0.2
		Shelikof Strait				
1470.2	1477.7	1465.5	1464.9	− 7.5*	+4.7*	+5.3*
1470.1	1477.7	1465.2	1465.1	− 7.6*	+4.9*	+5.0*
1474.0	1488.5	1471.6	1466.6	−14.5*	+2.4*	+7.4*
1466.0	1488.9	1470.7	1466.4	−22.9	−4.7	−0.4
1468.2	1477.2	1466.2	1464.9	− 9.0*	+2.0*	+3.3*
1473.0	1477.2	1465.5	1464.5	− 4.2*	+7.5*	+8.5*
1467.8	1488.9	1474.7	1469.1	−21.1	−6.9	−1.3
				Av. −12.4	Av. +1.4	Av. +4.0
		Oregon and Washington				
1477.8	1495.9	1487.6	1479.0	−18.1	−9.8	−1.2
1475.4	1495.9	1480.7	1475.4	−20.5	−5.3	0.0
1478.5	1497.8	1484.7	1477.3	−19.3	−6.2	+1.2
1482.1	1497.8	1486.0	1478.7	−15.7	−3.9	+3.4
1475.2	1497.8	1483.7	1476.2	−22.6	−8.5	−1.0
1475.3	1497.8	1483.3	1475.7	−22.5	−8.0	−0.4
1476.7	1497.8	1480.3	1475.4	−21.1	−3.6	+1.3
1476.1	1497.8	1480.1	1475.7	−21.7	−4.0	+0.4
1477.4	1497.8	1483.7	1476.1	−20.4	−6.3	+1.3
1476.4	1497.8	1484.7	1477.4	−21.4	−8.3	−1.0
				Av. −20.3	Av. −6.3	Av. +0.4

FIGURE 5

Comparison of experimental velocities with theoretical velocities for various paths

(Continued on next page)

FIGURE 5 (*Concluded*)

EXPERIMENTAL VELOCITY (E)	SURFACE VELOCITY (S)	MEAN VELOCITY (M)	BOTTOM VELOCITY (B)	E − S	E − M	E − B
		Northern California				
1479.9	1489.7	1482.8	1479.0	− 9.8	−2.9	+0.9
1477.2	1489.7	1481.9	1478.7	−12.5	−4.7	−1.5
1479.2	1489.7	1482.4	1479.1	−10.5	−3.2	+0.1
1477.2	1489.7	1481.6	1478.6	−12.5	−4.4	−1.4
1480.9	1489.7	1481.7	1478.5	− 8.9	−0.8	+2.4
1479.3	1489.7	1481.3	1478.5	−10.4	−2.0	+0.8
1478.9	1489.7	1481.4	1478.3	−10.8	−2.5	+0.6
1481.1	1489.7	1481.7	1478.3	− 8.6	−0.6	+2.8
1490.1	1497.3	1492.0	1489.3	− 7.2	−1.9	+0.8
1489.1	1494.5	1492.0	1491.2	− 5.4	−2.9	−2.1
1490.6	1494.5	1493.0	1491.7	− 3.9	−2.4	−1.1
1483.7	1497.3	1490.5	1486.7	−13.6	−6.8	−3.0
1488.0	1497.3	1490.5	1486.2	− 9.3	−2.5	+1.8
1489.2	1498.8	1491.7	1489.4	− 9.6	−2.5	−0.2
1485.2	1499.2	1488.1	1485.0	−14.0	−2.9	+0.2
1480.2	1489.0	1485.0	1480.2	− 8.8	−4.8	0.0
1479.7	1489.0	1484.2	1480.4	− 9.3	−4.5	−0.7
1480.9	1489.0	1483.0	1480.3	− 8.0	−2.1	+0.9
1481.0	1489.0	1483.8	1480.0	− 8.1	−2.8	+0.7
1478.2	1489.0	1483.4	1479.9	−10.8	−5.2	−1.7
1479.0	1489.0	1483.8	1480.7	−10.0	−4.8	−1.7
1493.8	1493.2	1487.4	1484.9	+ 0.6*	+6.4*	+8.9*
1487.8	1493.2	1487.9	1485.8	− 5.4	−0.1	+2.0
1490.4	1493.2	1487.4	1485.5	− 2.8*	+3.0*	+4.9*
1486.2	1493.2	1489.6	1486.7	− 7.0	−3.4	−0.5
1489.0	1492.1	1487.4	1483.1	− 3.1*	+1.6*	+5.9*
1489.0	1492.1	1488.8	1485.2	− 3.1*	+0.2*	+3.8*
				Av. − 8.3	Av. −2.2	Av. +0.9
			Grand Average	−12.3	−3.2	+1.0

result is a most extraordinary correspondence between measured velocities and theoretical bottom velocities.

Figure 6 gives the average differences between experimental and theoretical velocities computed from best observations (this excludes all thoscomparisons marked with an asterisk in Fig. 5) and from observations made in localities where the temperature gradient is high, such as exists in Alaska and off the coasts of Washington and Oregon. The latter is, after all, the critical test as far as determining the path of the effective sound energy, for where the temperature difference between surface and bottom is slight, the experimental velocity would be found to agree with almost any theory of sound wave propagation.

This remarkably close agreement between measured velocities and theoretical velocities based on bottom temperatures, would seem to indi-

cate that the peak of the energy reaching the hydrophone has come not by way of the shorter straight line path where the velocity is greater, but by way of the more circuitous path of the bottom layers of water where the velocity is actually less. How is this seeming contradiction to be explained? It has been suggested that it is primarily a question of temperature; that a given amount of energy will travel farther in cold water than in warm water, and though the velocity of sound is greater near the surface, the energy is used up quicker and over long distances fails to reach the hydrophone. Hence the only record we get is of the energy that has come by way of the colder bottom layers.

While there is considerable evidence to be found in support of this theory, both in the present study and in our experiences with horizontal trans-

	SURFACE	MEAN	BOTTOM
From all observations (51)	−12.3	−3.2	+1.0
From best observations (42)	−13.7	−4.6	+0.01
From observations in localities of high temperature gradient (17)	−18.7	−6.7	+0.15

FIGURE 6

Average differences (in meters per second) between experimental velocities and theoretical velocities for assumed paths.

mission on the south Atlantic coast, I believe it is too early to formulate a definite theory regarding the behavior of the sound wave. It will be time enough to consider these possibilities when we have supplemented our present data with experimental work carried out along certain lines which the investigation has shown is urgently needed.

For the present, the important thing is that a practical working relation has been established between experimental and theoretical velocities that has enabled us to adopt a definite policy for the work on Georges Bank. In addition, the study has shown that any assumption that the effective sound wave travels along the surface or close to the surface is wholly untenable. Other than that the investigation should be considered in the nature of a preliminary finding and as laying the foundation for a thorough and comprehensive study, both in the field and in the office, of the whole subject of sound transmission in all its ramifications.

A METRIC FOR THE SPACE OF MEASURABLE FUNCTIONS

By Andrew C. Berry*

Department of Mathematics, Princeton University

Communicated June 30, 1931

Fréchet[1] cites a metric for the space of measurable functions and shows that the space is complete with respect to this metric. He notes, however, that the metric does not satisfy the triangle inequality, and is not homogeneous. The present paper furnishes a metric which (Theorem 4) satisfies the triangle inequality, which (Theorem 5) possesses quasi-homogeneity, and for which (Theorem 6) the space is complete.

Definition. *A complex-valued, Lebesgue measurable function $f(x)$ of the single real variable x, $-\infty < x < \infty$, shall be said to be an element of the space M if there exists, as a finite number,*

$l(f)$ = greatest lower bound of real numbers ϵ such that

$$\text{meas}\left\{\,|f(x)| > \epsilon\,\right\} \leqq \epsilon.$$

Theorem 1. *A necessary and sufficient condition that $\lambda = l(f)$ is that*

$$\text{meas}\left\{\,|f(x)| > \lambda\,\right\} \leqq \lambda,$$

$$\text{meas}\left\{\,|f(x)| > \lambda - \delta\right\} > \lambda - \delta, \text{ for each } \delta > 0.$$

Proof of Necessity.—Let $\lambda = l(f)$. By definition, the second requirement is met. Again by definition,

$$\text{meas}\left\{\,|f(x)| > l(f) + \epsilon\,\right\} \leqq l(f) + \epsilon \text{ for each } \epsilon > 0.$$

As $\epsilon \longrightarrow 0$, the set $\left\{\,|f(x)| > l(f) + \epsilon\,\right\}$ *expands* toward the limit set $\left\{\,|f(x)| > l(f)\right\}$ and so lim meas = meas lim. Thus,

$$\text{meas}\left\{\,|f(x)| > l(f)\,\right\} \leqq l(f),$$

and the first requirement is met.

Proof of Sufficiency.—Let λ satisfy the two requirements. Then, by definition,

$$\lambda - \delta < l(f) \leqq \lambda$$

for each $\delta > 0$, which is possible only if $\lambda = l(f)$.

Theorem 2. $l(f) \geqq 0$.

Proof.—The inequality $\text{meas}\left\{\,|f(x)| > \epsilon\,\right\} \leqq \epsilon$ can be satisfied only by non-negative values of ϵ since measure is a non-negative quantity.

Theorem 3. *A necessary and sufficient condition that $l(f) = 0$ is that $f(x) = 0$ almost everywhere.*

Proof.—By Theorems 1 and 2, a necessary and sufficient condition

that $l(f) = 0$ is that meas $\{ |f(x)| > 0 \} = 0$, which is equivalent to the statement that $f(x) = 0$ almost everywhere.

THEOREM 4. $l(f+g) \leqq l(f) + l(g)$.

Proof.—If x is such that

$$|f(x) + g(x)| > l(f) + l(g),$$

then x must be such that at least one of the inequalities

$$|f(x)| > l(f), \quad |g(x)| > l(g)$$

is satisfied. Hence,

$$\text{meas } \{ |f(x) + g(x)| > l(f) + l(g) \} \leqq \text{meas } \{ |f(x)| > l(f) \} + \text{meas } \{ |g(x)| > l(g) \} \leqq l(f) + l(g),$$

the last inequality being a consequence of Theorem 1. It follows, by definition, that

$$l(f+g) \leqq l(f) + l(g).$$

THEOREM 5. *If* $|a| \geqq 1$, *then* $l(af) \leqq |a| l(f)$.

Proof.—By Theorem 1, when $|a| \geqq 1$,

$$\text{meas } \{ |af(x)| > |a| l(f) \} = \text{meas } \{ |f(x)| > l(f) \} \leqq l(f) \leqq |a| l(f).$$

Hence, by definition, $l(af) \leqq |a| l(f)$ when $|a| \geqq 1$.

THEOREM 6.[2] *If* $\{ f_n(x) \}$ *is a sequence of elements of* M *such that* $\lim_{m,n \to \infty} l(f_m - f_n) = 0$, *then there exists an element* $f(x)$ *of* M *such that*

$$\lim_{n \to \infty} l(f - f_n) = 0.$$

Proof.—We can determine a sub-sequence $\{ f_{n_k}(x) \}$ such that $l(f_m - f_n) \leqq 2^{-k}$ when $m \geqq n_k$ and $n \geqq n_k$, $k = 1, 2, \ldots$.

In particular, by Theorem 1,

$$\text{meas } \{ |f_{n_{k+1}}(x) - f_{n_k}(x)| > 2^{-k} \} \leqq 2^{-k}.$$

Denote by E_k the set

$$\sum_{j=k}^{\infty} \left\{ |f_{n_{j+1}}(x) - f_{n_j}(x)| > 2^{-j} \right\}.$$

Observe that each E_k includes the corresponding E_{k+1} and that

$$\text{meas } E_k \leqq 2^{-k+1}.$$

The series

$$f_{n_1}(x) + \{ f_{n_2}(x) - f_{n_1}(x) \} + \{ f_{n_3}(x) - f_{n_2}(x) \} + \ldots$$

converges uniformly over the complement of any E_k and so converges for almost all x to some limit function $f(x)$. On the complement of E_k,

$$| f(x) - f_{n_k}(x) | \leqq 2^{-k+1}.$$

Thus

$$\text{meas } \{ | f(x) - f_{n_k}(x) | > 2^{-k+1} \} \leqq \text{meas } E_k \leqq 2^{-k+1},$$

which implies that

$$l(f - f_{n_k}) \leqq 2^{-k+1}.$$

Finally, by Theorem 4,

$$\limsup_{n \to \infty} l(f - f_n) \leqq l(f - f_{n_k}) + \limsup_{n \to \infty} l(f_n - f_{n_k}) \leqq 2^{-k+1} + 2^{-k}, \qquad k = 1, 2, \ldots,$$

and this is possible only if

$$\lim_{n \to \infty} l(f - f_n) = 0.$$

We need only add the remark that since $f - f_{n_1}$ is in M and since f_{n_1} is in M, Theorem 4 proves that f is also in M.

NOTE 1. If, for each C, $O < C < \infty$, we define a space M_C as the set of all measurable functions $f(x)$ for which there exists, as a finite number,

$$l_C(f) = \text{greatest lower bound of real numbers } \epsilon \text{ such that}$$
$$\text{meas } \{ | f(x) | > \epsilon \} \leqq C\epsilon,$$

then, since we readily calculate that

$$l_{C'}(f) \leqq l_C(f) \leqq \frac{C'}{C} l_{C'}(f) \text{ when } o < C < C',$$

we see that all the spaces M_C coincide with the space M as far as their elements are concerned, and that convergence in any of the spaces implies convergence in all the other spaces.

NOTE 2. Each integrability space L_p is contained in M in the sense that if $f(x)$ is an element of some L_p it is also an element of M, and that convergence in L_p implies convergence in M. For, if we use the usual definitions

$$l^{(p)}(f) = \left\{ \int_{-\infty}^{\infty} | f(x) |^p \, dx \right\}^{\frac{1}{p}}, \quad 1 \leqq p < \infty,$$

$l^{(\infty)}(f)$ = greatest lower bound of numbers ϵ such that

$$\text{meas } \{ | f(x) | > \epsilon \} = 0,$$

then we find readily that

$$\begin{cases} l(f) \leqq [l^{(p)}(f)]^{\frac{p}{p+1}}, & 1 \leqq p < \infty, \\ l(f) \leqq l^{(\infty)}(f). \end{cases}$$

* NATIONAL RESEARCH FELLOW.

[1] M. Fréchet, "Sur divers modes de convergence d'une suite de fonctions d'une variable," *Bull. Calcutta Math. Soc.*, **11**, 187–206 (1921); "Les espaces abstraits," Paris, Gauthier-Villars, p. 91 (1928).

[2] The proof here is essentially the standard proof in the theory of convergence in measure. See, for instance, F. Riesz, *Compt. Rend.*, **144**, 615–619 (1907); E. Fischer, *Compt. Rend.*, **144**, 1022–1024 (1907); H. Weyl, *Math. Annal.*, **67**, 225–245 (1909).

THE EULER EQUATIONS OF PROBLEMS OF THE CALCULUS OF VARIATIONS WITH PRESCRIBED TRANSVERSALITY CONDITIONS

BY LINCOLN LA PAZ

DEPARTMENT OF MATHEMATICS, OHIO STATE UNIVERSITY

Communicated June 30, 1931

If n functions $T_i(x, y_1, \ldots, y_n, y_1', \ldots, y_n')$ are of class C''^1 in a region S of $(x, y_1, \ldots, y_n, y_1', \ldots, y_n')$-values and in S satisfy the inequalities

$$D = \left| \partial T_i / \partial y_j' \right| \neq 0 \quad (i, j = 1, \ldots, n), \quad 1 + y_\mu' T_\mu \neq 0 \qquad (1)$$

and the $n(n-1)/2$ relations

$$\Gamma_{mk} = 0 \ (m, k = 1, \ldots, n, m < k), \qquad (2)$$

where

$$\Gamma_{mk} = y_\mu'(T_m \partial T_k / \partial y_\mu' - T_k \partial T_m / \partial y_\mu') + (\partial T_k / \partial y_m' - \partial T_m / \partial y_k'), \qquad (3)$$

then there always exists a *non-singular* problem of minimizing an integral

$$I = \int_{x_1}^{x_2} f(x, y_1, \ldots, y_n, y_1', \ldots, y_n')dx \qquad (4)$$

with an integrand function f of class C''' which has the functions T_i as its *transversality coefficients*.[2] In (1) and (3) as elsewhere in this paper a repeated Greek letter is an umbral index indicating a summation with the range 1 to n.

The most general such problem (4) has an integrand function

$$f = ge^h \qquad (5)$$

where $g(x, y_1, \ldots, y_n)$ is a function different from zero and of class C''' in the $(x, y_1, \ldots, y_n)$-projection of S but otherwise arbitrary, and where

$$h = \int_0^{y'} T_\mu dy_\mu' / (1 + y_\mu' T_\mu) \qquad (6)$$

is a line integral which in a space of suitably simple connectivity properties

can be shown to be independent of the path of integration in view of relations (2).[3] In (6) the symbols 0 and y' used as limits represent sets of n quantities $(0, \ldots, 0)$ and $(y'_1, \ldots, y'_n)$.

It is clearly no restriction to assume $g > 0$, since a problem for which $g < 0$ can always be transformed into a problem for which $g > 0$ by requiring that the integral I in (4) be maximized instead of minimized. Set

$$l = \log g, \qquad H = h + l. \tag{7}$$

Then the Euler equations of (4) formed with the integrand function (5) are

$$\frac{d}{dx}\left(\frac{\partial e^H}{\partial y'_i}\right) - \left(\frac{\partial e^H}{\partial y_i}\right) = 0 \qquad (i = 1, \ldots, n). \tag{8}$$

In view of relations (6) and (7), the system (8) reduces at once to

$$[T_{iy'_\gamma} + y'_\beta(T_\beta T_{iy'_\gamma} - T_i T_{\beta y'_\gamma})]y''_\gamma = M_i \ (i = 1, \ldots, n) \tag{9}$$

where

$$\begin{aligned} M_i = {} & [(1 + y'_\mu T_\mu)H_{y_i} - T_i y'_\mu H_{y_\mu} - T_i H_x](1 + y'_\mu T_\mu) \\ & + [T_i y'_\beta T_{\beta y_\gamma} - (1 + y'_\mu T_\mu)T_{iy_\gamma}]y'_\gamma \\ & + T_i y'_\beta T_{\beta x} - (1 + y'_\mu T_\mu)T_{ix}. \end{aligned} \tag{10}$$

To obtain a system of differential equations in normal form from these equations the determinant of the coefficients of the second derivatives must be evaluated. This determinant

$$d = \left| T_{iy'_j} + y'_\beta(T_\beta T_{iy'_j} - T_i T_{\beta y'_j}) \right| (i, j = 1, \ldots, n), \tag{11}$$

is recognized as the product of the determinants

$$\Delta = \left| \delta^i_j \ (1 + y'_\mu T_\mu) - \delta^j_\mu y'_\mu T_i \right| \tag{12}$$

and $D = \left| \partial T_i / \partial y'_j \right|$. To evaluate Δ we note that

$$\left| \delta^i_j + y'_i T_j \right| \cdot \Delta = \left| \delta^i_j \ (1 + y'_\mu T_\mu) \right| = (1 + y'_\mu T_\mu)^n. \tag{13}$$

It is not difficult to show that $\left| \delta^i_j + y'_i T_j \right| = (1 + y'_\mu T_\mu)$. Hence (13) gives $\Delta = (1 + y'_\mu T_\mu)^{n-1}$.

Consequently $d = D \cdot (1 + y'_\mu T_\mu)^{n-1}$ and therefore in the region S the second derivatives y''_k can be obtained in the form

$$y''_k = d_k / D \cdot (1 + y'_\mu T_\mu)^{n-1}, \tag{14}$$

where d_k is the determinant d with the element d_{ik} replaced by the quantity M_i in (10) for $i = l, \ldots, n$. The determinant which arises when the element in the ith row and kth column of d is replaced by M_i is found to have the value

$$d_k = D_k \cdot (1 + y'_\mu T_\mu)^{n-1}, \tag{15}$$

where D_k is the determinant D with the element $\partial T_i / \partial y'_k$ replaced for $i = 1, \ldots, n$ by the quantity

$$N_i = (H_{y_i} - T_i H_x)(1 + y'_\mu T_\mu) - (T_{ix} + T_{iy_\mu} y'_\mu). \tag{16}$$

Hence the system (14) reduces to the compact system of equations in normal form described in the following theorem:

THEOREM 1. *For problems* (4) *with a prescribed set of transversality coefficients* T_i, *the Euler equations may be put in the normal form*

$$y_k'' = D_k/D \quad (k = 1, \ldots, n). \tag{17}$$

In (17) D *is the Jacobian of the* T_i *with respect to the* y_j' *and* D_k *is the determinant obtained from* D *by replacing the element* $\partial T_i/\partial y_k'$ *in* D *by the quantity* N_i *in* (16) *for* $i = 1, \ldots, n$.

For the special case in which transversality, in the sense of the calculus of variations, is to be equivalent to orthogonality the transversality coefficients T_i must have the values

$$T_i(x, y_1, \ldots, y_n, y_1', \ldots, y_n') = y_i' \quad (i = 1, \ldots, n). \tag{18}$$

When these values for T_i are employed the system (17) reduces to

$$y_k'' = (\partial l/\partial y_k - y_k'\partial l/\partial x)(1 + y_\mu' y_\mu') \quad (k = 1, \ldots, n). \tag{19}$$

An obvious necessary and sufficient condition for (19) to reduce to the system

$$y_k'' = 0 \quad (k = 1, \ldots, n) \tag{20}$$

is that l be a constant. Hence, since two integrals (4) for which the integrand functions differ only by a constant factor are not regarded as distinct, we reach the following theorem:

THEOREM 2. *The only problem of the calculus of variations in space of* $(n + 1)$*-dimensions for which transversality is orthogonality and for which the extremals are the straight lines satisfying the system* (20), *is the problem of minimizing the generalized length integral*

$$I = \int_{x_1}^{x_2} (1 + y_1'^2 + \ldots + y_n'^2)^{\frac{1}{2}} dx. \tag{21}$$

In case $n = 2$ and we set $y = y_1$, $z = y_2$ the system (19) reduces to the pair of equations

$$\begin{aligned} y'' &= (\partial l/\partial y - y'\partial l/\partial x)(1 + y'^2 + z'^2), \\ z'' &= (\partial l/\partial z - z'\partial l/\partial x)(1 + y'^2 + z'^2). \end{aligned} \tag{22}$$

The distinctive form of the equations (22) has been employed[4] as a criterion for a *natural family*, that is to say, a four-parameter family of curves in three-space which are identifiable with the extremal system of a problem of minimizing an integral of the *special* form

$$I = \int_{x_1}^{x_2} F(x, y, z)(1 + y'^2 + z'^2)^{\frac{1}{2}} dx. \tag{23}$$

In this same sense the system (17) may be regarded as a characteristic normal form of the Euler equations of a problem of minimizing the *general* integral (4) in space of $(n + 1)$-dimensions. For, as has been shown, in

terms of the transversality coefficients $(\partial f/\partial y_i')/(f - y_\mu'\partial f/\partial y_\mu')$ $(i = 1, \ldots, n)$, of a problem (4) the Euler system can always be reduced to the normal form (17); and, conversely, if a system of differential equations

$$y_k'' = F_k(x, y_1, \ldots, y_n, y_1', \ldots, y_n') \quad (k = 1, \ldots, n), \tag{24}$$

permits of representation in the special form (17) then it is a normal Euler system of a problem (4) for which the transversality coefficients are the quantities T_i and the integrand function is $f = e^H$. Thus a necessary and sufficient condition for the differential equations (24) to be a system of normal Euler equations of a problem (4) in the calculus of variations is that there exist n functions $T_i(x, y_1, \ldots, y_n, y_1', \ldots, y_n')$ $(i = 1, \ldots, n)$, possessing the properties of transversality coefficients, and a function $H(x, y_1, \ldots, y_n, y_1', \ldots, y_n')$ which together satisfy the system of partial differential equations

$$\partial H/\partial y_k' = T_k/(1 + y_\mu' T_\mu), \; F_k = D_k/D \quad (k = 1, \ldots, n). \tag{25}$$

The integrability conditions for the sub-system (25_1) are the $n(n-1)/2$ relations

$$R_{ij} = 0 \quad (i, j = 1, \ldots, n, \; i < j), \tag{26}$$

where

$$\begin{aligned} R_{mk} = y_\mu'[T_m \partial T_\mu/\partial y_k' - T_k \partial T_\mu/\partial y_m' + T_\mu(\partial T_k/\partial y_m' \\ - \partial T_m/\partial y_k')] + (\partial T_k/\partial y_m' - \partial T_m/\partial y_k'). \end{aligned} \tag{27}$$

If we note that in terms of the R_{ij} the quantities Γ_{mk} in (3) are expressible in the form

$$\Gamma_{mk} = (R_{mk} - y_\mu' T_k R_{\mu m} - y_\mu' T_m R_{k\mu})/(1 + y_\mu' T_\mu), \tag{28}$$

and that in view of (25_2) the quantities F_j are the unique solutions obtained by Cramer's Rule from the system of equations

$$(\partial T_j/\partial y_\mu')F_\mu = N_j \quad (j = 1, \ldots, n), \tag{29}$$

the statements of the last paragraph may be formulated in the following theorem:

THEOREM 3. *A necessary and sufficient condition for the equations* (24) *to be a system of normal Euler equations of a non-singular problem* (4) *of the calculus of variations with integrand function f of class C''' is that there exist n functions* $T_i(x, y_1, \ldots, y_n, y_1', \ldots, y_n')$ *of class C'' satisfying the inequalities* (1) *and a function* $H(x, y_1, \ldots, y_n, y_1', \ldots, y_n')$ *such that the following 2n equations*

$$\partial H/\partial y_k' = T_k/(1 + y_\mu' T_\mu), \; (\partial T_k/\partial y_\alpha')F_\alpha = N_k \; (k = 1, \ldots, n), \tag{30}$$

where N_k *is the function defined in* (16), *are satisfied identically in* $x, y_1, \ldots, y_n, y_1', \ldots, y_n'$.

Since

$$R_{mk} = (1 + y'_\mu T_\mu)\Gamma_{mk} + y'_\mu T_k \Gamma_{\mu m} + y'_\mu T_m \Gamma_{k\mu}, \qquad (31)$$

the compatibility of the sub-system (30_1) is insured by the relations (2) which n functions T_i must satisfy if they are to be transversality coefficients, but the determination of the integrability conditions for the entire system (30) appears to be difficult and will be considered in a later paper.

As an immediate corollary to Theorem 3 we have

THEOREM 4. *A necessary and sufficient condition for the integral curves of the system* (24) *to be the extremals of a problem* (4) *for which transversality is orthogonality is that there exists a single function* $l(x, y_1, \ldots, y_n)$ *satisfying the system of partial differential equations*

$$F_k = (\partial l/\partial y_k - y'_k \partial l/\partial x)(1 + y'_\mu y'_\mu) \quad (k = 1, \ldots, n). \qquad (32)$$

[1] The term class is used in the sense of Bolza. Compare *Vorlesungen über Variationsrechnung*, p. 13.

[2] La Paz, L., *Bull. Amer. Math. Soc.*, **36**, p. 674, 1930.

[3] In this connection a function f of the form (5) was first obtained by Rawles, *Trans. Amer. Math. Soc.*, **30**, p. 778, 1928. A proof that the formula (5) actually furnishes the most general integrand function of a non-singular problem with a specified transversality was first given by the writer, loc. cit., p. 680.

[4] Kasner, E., "The Theorem of Thompson and Tait and Natural Families of Trajectories," *Trans. Amer. Math. Soc.*, **11**, pp. 121–140, 1910; "Differential-Geometric Aspects of Dynamics," *Princeton Colloquium Lectures*, § 38, 1912.

THEOREMS RELATING TO THE HISTORY OF MATHEMATICS

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated June 22, 1931

Attention has frequently been directed to evidences of the dominance of the group concept in the development of mathematics, and in the second edition (1927) of his well-known *Theorie der Gruppen von endlicher Ordnung* A. Speiser emphasized symmetry in ancient ornaments as an evidence of a prehistoric group theory extending back to about 1500 B. C. It may be desirable to note here a few evidences of a lack of dominance of the group concept in developments closely related thereto in order to obtain a truer picture of the actual influence of this concept. The very late explicit formulation of the group concept is a striking instance in the history of mathematics of extreme slowness in coördinating related developments and it deserves emphasis on this account.

Special arithmetic and geometric series appear in the writings of the

ancient Babylonians and of the ancient Egyptians. If such a series is supposed to be extended infinitely in both directions then a necessary and sufficient condition that it constitutes a group is that it contains the sum of at least one pair of its terms, if it is arithmetic, and that it contains the product of at least one pair of its terms, if it is geometric. We have therefore here instances relating to some of the earliest developments of mathematics where the existence of a group can be established by a single elementary trial. It would be very interesting to find examples where this existence was actually established in this way, if such examples can be found in the mathematical literature. The fact that it was not emphasized exhibits a lack of the dominance of the group concept in this case.

The slowness with which zero and negative numbers were adopted as members of our system of numbers constitutes another evidence of the lack of dominance of the group concept, since the integers became a group when they are combined by addition only after these numbers were adopted. On the other hand, the positive rational numbers constitute a group when they are combined by multiplication and the number concept was implicitly extended very early so as to include all of these numbers. As the zero is the identity of the addition group its comparatively late appearance as an actual number constitutes one of the most interesting examples of the lack of dominance of the group concept in early mathematical developments, especially since it relates to the foundation thereof as addition is one of the earliest mathematical operations.

Another forceful example of this lack of dominance is furnished by the theorems relating to regular solids which appear in Euclid's *Elements*. These theorems do not emphasize the properties of the special groups belonging to these solids as might have been expected if an actual intensive prehistoric group theory is assumed to have then existed. On the contrary, even the group of such a simple figure as the square does not appear to have been explicitly noted before the nineteenth century. Hence it results that the early mathematical developments were not as intensively affected by the group concept as might be inferred from some of the statements relating thereto. The intensive period began in the latter half of the eighteenth century and became continually more pronounced during the nineteenth and the beginning of the twentieth century. If we assume earlier periods of group theory developments it should be emphasized that these developments were then of a much less intensive type and fade into mere traces.

If one term of an arithmetic series, extended infinitely in both directions, is a multiple of the common difference of the series then every term of the series has this property, and if one term of such a geometric series is a power of the common ratio then this is also true of every other term thereof. That is, *a necessary and sufficient condition that an arithmetic or a geometric*

series, extended infinitely in both directions, constitutes a group with respect to addition or multiplication, respectively, is that it involves at least one term which is a multiple of its common difference or a power of its common ratio, respectively. A (1,1) correspondence in which the identity of one of these series corresponds to the identity of the other can therefore be established in such a way that to the product of any two terms of one series there corresponds the sum of the corresponding terms of the other, and vice versa. Such a correspondence is of fundamental importance in the history of logarithms and exhibits contact between the group concept and the theory of logarithms. The trigonometric approach to logarithms was developed in a subject known as prosthaphaeresis, which also connects the operations of addition and multiplication in an explicit form and was developed on this account.

It is a singular fact that in the (1,1) correspondence between the two series employed by John Napier in the construction of his so-called logarithmic tables the identities of the related groups were not placed in correspondence and hence to the product of two terms of his geometric series there corresponds the sum of the two corresponding terms of the related arithmetic series plus a constant. Somewhat similar remarks apply to the relation which Archimedes noted between a partial arithmetic series and the corresponding terms of a geometric series, except that at the time of Archimedes only half-groups with respect to an arithmetic series could be considered since only positive numbers were then employed. The works of Archimedes and Napier can be more fully understood in the light of such group theory theorems even if these theorems were unknown at the time when these works were written and did not explicitly affect them. The theorem that *every arithmetic series and every geometric series becomes a group when the appropriate terms are interpolated and exterpolated* has obvious contact here.

The rôle played by half-groups in the history of elementary mathematics is perhaps of equal importance to that played by the actual groups therein. In addition to the half-group noted in the preceding paragraph the half-group composed of the positive unit fractions when they are combined by multiplication is of fundamental historical importance in view of the prominence of these fractions in the works of the ancient Egyptians and others. Attention was recently directed to the following theorem relating to these fractions: *An arbitrary rational number can be expressed as the sum of the reciprocals of a finite set of distinct integers having the same sign and such that the smallest absolute value of an integer of the set is arbitrarily large and that the number of these integers in every set exceeds any assigned natural number.*[1] Hence it results that a definition of the term rational number can be based on the property noted in this theorem. This seems not to have been done hitherto notwithstanding its important historical connections.

A striking difference between the histories of analysis and geometry is that while the latter was based in part upon a system of postulates by the ancient Greeks the former was developed without any explicit system of postulates up to recent times. A mathematical theorem relating thereto is that the angle formed by a curve and its tangent line is an actually infinitely small magnitude and hence does not obey the so-called Archimedean postulate. By making the implicit assumption that all rectilinear angles obey this postulate Euclid proved that this angle is smaller than any rectilinear angle. This proof points to an interest on the part of the ancient Greeks in angles which are not rectilinear and throws some light on the formulation by them of the postulate that all right angles are equal. In a comparatively concrete subject like geometry the need of a system of postulates naturally became apparent earlier than in the more abstract subject of analysis. At any rate, there is here an interesting disparity as regards the use of postulates in the development of mathematics.

Just as points in geometry correspond to the existence of certain subgroups of the group of movements so the concept of natural number corresponds to the group of permutations of the collection representing this number when the individuals thereof are permuted in every possible way. The natural number n may therefore be regarded as an invariant under the symmetric permutation group of degree n. This symmetric group may thus be related with one of the earliest concepts of mathematics even if its relation thereto was naturally not explicitly formulated in ancient times. In Webster's *New International Dictionary* (1927) under the term "distribution" it is stated that "the arrangement of objects in classes called *parcels* when the order of the objects is indifferent, and *groups* when it is essential" represents a distribution. This definition of the term group does not seem to be in accord with the facts noted above and may deserve notice here in view of the fact that the work in which it appears is so widely used, and improvements frequently depend upon persistent demand therefor.

[1] G. A. Miller, *Amer. Math. Monthly*, **38**, 194, 1931.

ON THE INTENSITY OF TOTAL SCATTERING OF X-RAYS BY GASES. I

BY Y. H. WOO

DEPARTMENT OF PHYSICS, NATIONAL TSING HUA UNIVERSITY, PEIPING, CHINA

Communicated June 22, 1931

In previous notes[1] a calculation has been made of the intensity of total scattering of x-rays by the atoms of a monatomic gas, and fair agreement is obtained with the experiments of Barrett[2] on the scattering of x-rays by helium and argon gases and of Scherrer and Staeger[3] on the scattering of CuKα radiation by mercury vapor. The theoretical formula for the scattering is derived on the basis of classical electrodynamics by A. H. Compton[4] and C. V. Raman[5] with the aid of a dynamic atomic model, in which the electrons are regarded as a gas distributed in an enclosure surrounding the nucleus. The intensity of the x-rays of wave-length λ scattered at an angle θ to a distance R is found to be

$$I_{\theta 1} = I_1 + I_2[1 + \gamma(1 - \cos\theta)]^{-3}. \qquad (1)$$

In this formula $\gamma = h/mc\lambda$,

$$I_1 = \frac{Ie^4(1 + \cos^2\theta)}{2R^2m^2C^4} F^2 \text{ and } I_2 = \left(\frac{Ie^4(1 + \cos^2\theta)}{2R^2m^2C^4} Z - \frac{F^2}{Z}\right),$$

where I is the intensity of the primary beam, F is the atomic structure factor and h, e, m, c and Z have their usual significance. The factor $[1 + \gamma(1 - \cos\theta)]^{-3}$ is introduced by Compton[4] to correct for the change of wave-length in accordance with the theoretical investigation on the Compton effect by Breit,[6] Dirac[7] and others. From classical consideration Raman[5] has in particular called attention to the incoherent or "fluctuating" character of the part represented by I_2 and emphasized the fundamental difference in physical nature between I_1 and I_2.

In considering the scattering of x-rays by a gas molecule consisting of several atoms, it is necessary to take account of the phases of the rays scattered by the different atoms according to the classical wave-principles. We shall assume that only the coherent radiation scattered by the different atoms in the molecule will interfere with each other, while the incoherent radiation, owing to its "fluctuating" character as emphasized by Raman,[5] will be simply added together. This is the fundamental viewpoint taken in this paper.

Now let us consider the scattering of the x-rays of wave-length λ by a gas molecule containing atoms 1, ..., i, ..., n arranged at fixed distances from each other and each being characterized by an atomic structure factor F_i. Then, taking the average over all possible orientations,

the intensity of the rays scattered at an angle θ to a distance R is given by

$$I_{\theta n} = I_e \sum_1^n \sum_1^n F_i F_j \frac{\sin X_{ij}}{X_{ij}} + I_e \sum_1^n \frac{Z_i - \frac{F_i^2}{Z_i}}{[1 + \gamma(1 - \cos\theta)]^3}. \qquad (2)$$

In this equation the double summation is to be taken twice over all the atoms in the molecule,

$$I_e = \frac{I_e^4(1 + \cos^2\theta)}{2R^2m^2C^4} \text{ and } X_{ij} = kS_{ij} \qquad (3)$$

where s_{ij} is the distance of any atom i from any other atom j and $k = (4\pi/\lambda)\sin\frac{\theta}{2}$. The first term occurring in the right-hand side of equation (3) is the well-known expression derived by Debye,[8] while the second term represents the incoherent intensity due to the scattering by all the atoms composing the molecule. As one of the simplest applications we take the case of diatomic molecules like H_2, O_2.... Since there exists only one function F, the scattered intensity is thus

$$I_{\theta 2} = 2I_e \left\{ \left(1 + \frac{\sin X}{X}\right) F^2 + \frac{Z - \frac{F^2}{Z}}{[1 + \gamma(1 - \cos\theta)]^3} \right\} \qquad (4)$$

where, as in equation (3), $x = ks$, s being the distance between the two atoms in the diatomic molecule. Considering the nature of the atomic structure factor and the dependence of the incoherent intensity on λ and θ, we can at once conclude, in the first place, that the classically predicted maxima and minima will usually not show up in the scattering intensity, secondly, that in every case excess scattering should be present at small scattering angles and thirdly, that for the scattering of comparatively hard x-rays the intensity at large angles will approximately follow the law

$$I_{\theta\Omega} = \frac{2I_eZ}{[1 + \gamma(1 - \cos\theta)]^3} \qquad (5)$$

as predicted by Breit,[6] Dirac[7] and others. In accordance with the recent determination of moment of inertia by Rasetti,[9] the values of s are taken to be equal to 1.1A and 1.21A for H_2 and O_2, respectively, and a calculation according to equation (4) is made of the scattering of MoKα rays (λ = 0.71A) by these gases. In each case the atomic structure factor F is approximately estimated from the atomic field of Thomas and Fermi. The results of this calculation are compared with the experimental data recently reported by Barrett[10] in figure 1, where the scattering per electron in arbitrary units is plotted against the scattering angle θ. The classical theory of J. J. Thomson is plotted as the broken curve marked C and

the equation (5) is the broken curve marked Q. It is seen that for the case of oxygen the agreement between theory and experiment is as good as can be expected. It is also obvious that the experimental results obtained with general radiation could not be fitted to the theoretical curve. The calculated results for hydrogen are shown in curve I, which indicates that, at θ equal 20°, the interference begins to set in, while in the range 20° to 90° the scattering follows equation (5) closely (the theoretical curve practically coincides with the curve marked Q in this range). This is exactly what Barrett concludes on the basis of his experiments.[11] Owing to the comparatively low accuracy of the small angle readings, Barrett considers the evidence for the existence of the excess scattering not decisive, yet, judging from Barrett's results, the presence of this effect seems to be proved beyond doubt. Since the exact shape of the intensity curve depends entirely on the atomic structure factor F and the value of the distance s, the agreement between theory and experiment, therefore, points to the conclusion that the values of s taken for H_2 and O_2 are at least in the right order of magnitude. This also shows that interference measurements of this type give a direct method for determining the interatomic distance inside a molecule, if the atomic structure factor could be accurately estimated.

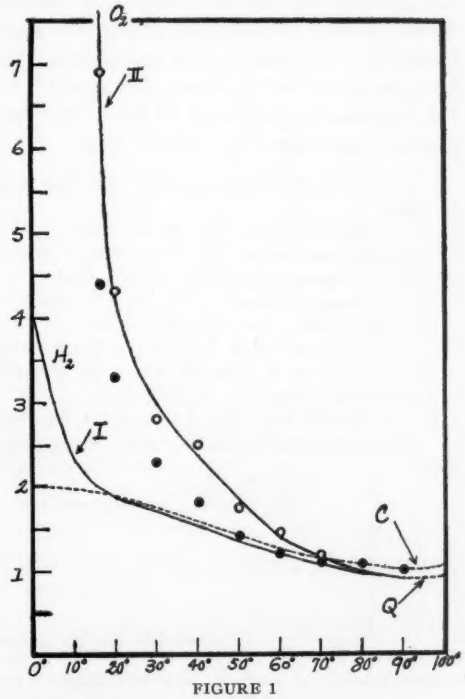


FIGURE 1

Barrett's Experiments { ○ From O_2, λ = 0.71A. ● From O_2, general radiation

In this connection we may mention the interesting work on the interference measurements with single molecules recently carried out by Debye and his co-workers.[11] Debye has interpreted the observed results according to equation (2) without the incoherent term. This is all right for the scattering of CuKα radiation by CCl_4. For the case of MoKα

rays the incoherent intensity should become appreciable at large angles. Unfortunately the microphotometer curve for this case has not been published, we cannot make any comparison here.

Finally it may be remarked that, according to the viewpoint put forward here, the intensity of the incoherent scattering per molecule should be independent of both the physical state and the temperature of the scattering material and the total scattering (incoherent) is simply proportional to the number of molecules effective in scattering. Thus we expect the incoherent scattering to play an important part in the scattering of x-rays by gases as well as by liquids and solids. This seems to be supported by the experiments performed to study the Compton effect. It is hoped to discuss these points in a future paper.

[1] Y. H. Woo, *Nature*, Oct. 4, 1930; *Proc. Nat. Acad. Sci.*, **16**, 418, 1930; and *Nature* in press.

[2] C. S. Barrett, *Phys. Rev.*, **32**, 22, 1928.

[3] Scherrer and Staeger, *Helv. Phys. Acta*, **1**, 518, 1928.

[4] A. H. Compton, *Phys. Rev.*, **35**, 925, 1930.

[5] C. V. Raman, *Indian J. Phys.*, **3**, 357, 1928.

[6] G. Breit, *Phys. Rev.*, **27**, 242, 1926.

[7] P. A. M. Dirac, *Proc. Roy. Soc.*, AIII, 405, 1926.

[8] P. Debye, *Ann. d. Phys.*, **46**, 809, 1915; cf. also *Phys. Zeits.*, **31**, 419, 1930.

[9] F. Rasetti, *Phys. Rev.*, **34**, 367, 1930.

[10] C. S. Barrett, *Proc. Nat. Acad. Sci.*, **14**, 20, 1928.

[11] Debye, Bewilogua and Ehrhardt, *Phys. Zeits.*, **30**, 84, 1929; *Ibid.*, **31**, 419, 1930.

ON THE INTENSITY OF TOTAL SCATTERING OF X-RAYS BY GASES. II

BY Y. H. WOO

DEPARTMENT OF PHYSICS, NATIONAL TSING HUA UNIVERSITY, PEIPING, CHINA

Communicated June 22, 1931

In a previous paper (to be referred to later as Part I) by the writer[1] a general theory of the intensity of total scattering of x-rays by the molecules of a polyatomic gas is developed on the assumption that, in considering the scattering of x-rays by a polyatomic molecule, only the coherent scattered radiation from the different atoms will interfere with each other, while the incoherent radiation will be simply added up. The theoretical intensity of the scattering of MoK_α rays by H_2 and O_2 is actually compared with the scattered intensity observed by Barrett[2] and the agreement seems to be satisfactory. In the present note it is intended (1) to make a detailed comparison of the theory given in Part I with Barrett's

experiments[3] on the scattering of x-rays by O_2 and N_2 for various wave-lengths, and (2) to discuss some points regarding to the factor $[1 + \gamma (1 - \cos \theta)]^{-3}$ introduced by A. H. Compton[4] to correct for the change of wave-length in the Compton effect.

(1) *Comparison of Theory with Experiment.*—As is shown in Part 1,

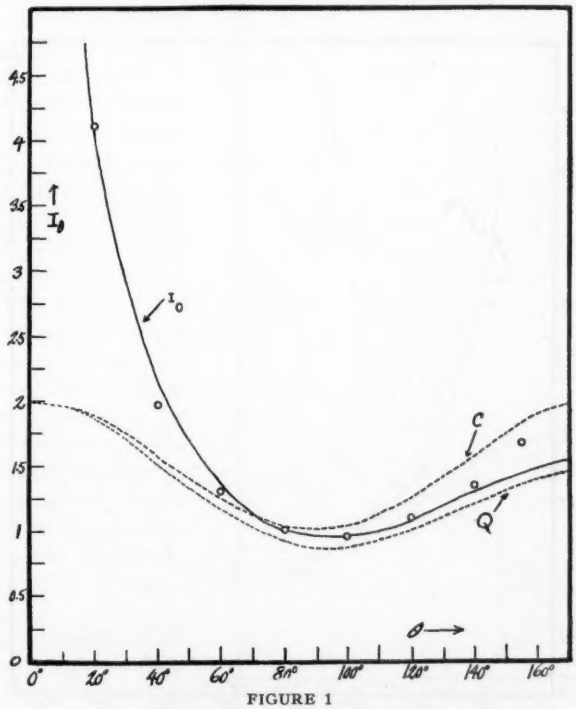


FIGURE 1

Barrett's results, ○ from O_2 for $\lambda = 0.48$A.

the intensity of the x-rays scattered at an angle θ to a distance R by a diatomic molecule like O_2 is given by

$$I_{\theta_2} = \frac{Ie^4(1 + \cos^2 \theta)}{2R^2m^2c^4} \left\{ 2\left(1 + \frac{\sin ks}{ks}\right) F^2 + 2 \frac{Z - \frac{F^2}{Z}}{[1 + \gamma(1 - \cos \theta)]^3} \right\},$$

where I is the intensity of the primary beam, F is the "atomic structure factor," s is the distance between the two atoms in the diatomic molecule, $k = (4\pi/\lambda)\sin {}^1/_2\theta$, $\gamma = h/mc\lambda$, and λ, Z, m, c and e have their usual significance. Taking the values of s to be equal to $1.21A$ and $1.10A$

for O_2 and N_2, respectively,[5] a calculation according to the method given in Part I is made of the intensity of total scattering of x-rays by these gases. The results of this calculation are compared with the experiments of Barrett[3] in figures 1 to 3, where the scattering per electron I_4, expressed in terms of the classical intensity of scattering by a single electron at $\theta = 90\lambda$, is plotted against the scattering angle θ. The curves I_0, II_0,

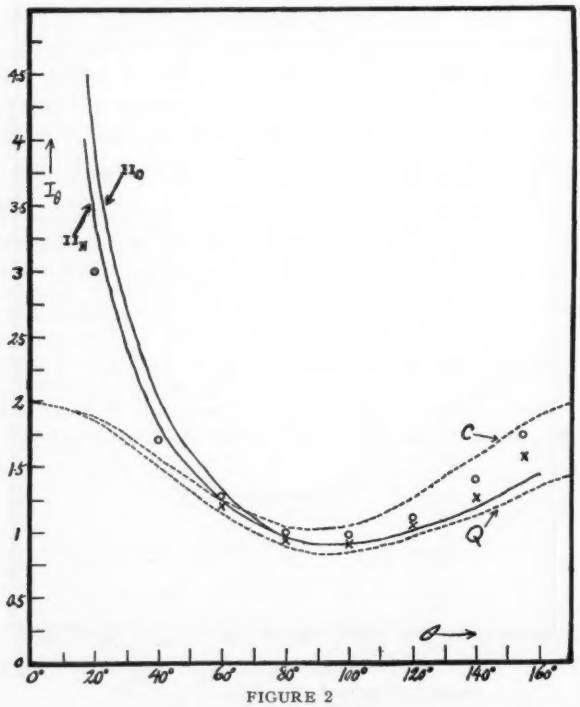


FIGURE 2

Barrett's results { × from O_2 ; ○ from N_2 } for $\lambda = 0.39$A.

and III_0 in figures 1 to 3 represent the scattering from O_2 for wave-lengths equal to 0.48A, 0.39A and 0.49 A, respectively, while the curves II_π and III_π in figures 2 and 3 represent the scattering from N_2 for wave-lengths equal to 0.39A and 0.49A, respectively. The theory of J. J. Thomson for the scattering from a single electron is plotted as the broken curve marked C and the quantum formula deduced by Breit,[6] Dirac[7] and others is the broken curve Q. Since the measurements of Barrett give the relative values of the scattering per electron for different angles, but not

absolute values, so in each case the experimental results have been multiplied by an arbitrary constant throughout. It is seen that the agreement between theory and experiment is in general satisfactory except for the readings at 20° and 155°. As pointed by Barrett himself,[3] because of the difficulty in obtaining the correction factors for these two angles the experimental data are less reliable than those for the other angles.

So far as the writer is aware, the measurements of Barrett seem to be the only ones available for a comparison with the present theory. Experiments on the scattering of MoKα rays by argon and carbon dioxide have been reported by Mark and Schocken,[8] but the accuracy of these experiments is not as great as that of Barrett's work. This comparison of theory and experiment shows, therefore, that accurate measurements on the intensity of total scattering of x-rays by gases and vapors for different wavelengths are much desired.

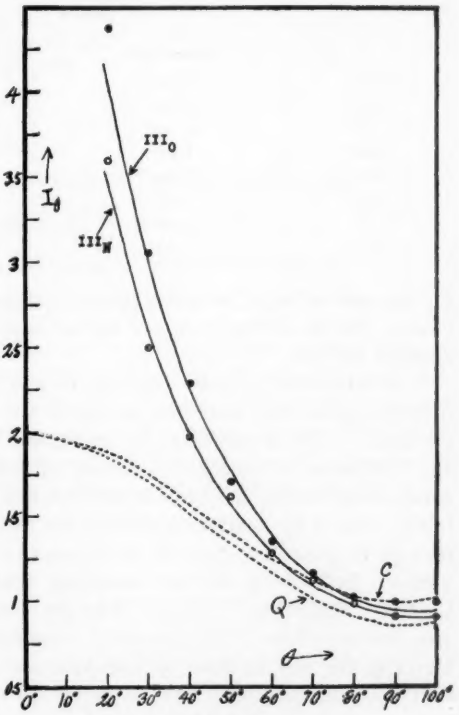


FIGURE 3

$\lambda = 0.49$A.

Barrett's Results { ● from O_2 ○ from N_2

(2) *On the Correction Factor* $[1+\gamma(1-\cos\theta)]^{-3}$. —It is true that the introduction of this factor by Compton is more or less arbitrary, but its significance is well illustrated by the following simple example. In table 1, numerical data of the scattering of x-rays of wavelength equal to 0.39A by argon are summarized.[9] The second column gives the scattering per electron I_θ, in the units adopted in the preceding section, calculated according to the formula with the correction factor. In the third column are given the values of the same quantity calculated without the correction factor. The experimental results in arbitrary units obtained by Barrett[3] are listed in the fourth column. It will be noticed that, while it

seems not possible to fit the results given in columns 3 and 4 with each other, the agreement between the theoretical values in column 2 and the observed data in column 4 is very satisfactory. This indicates that the correction factor, obtained from the consideration of quantum theory, refines the classical electrodynamical result so as to enable it to hold good

TABLE 1

θ	I_θ (CORRECTED)	I_θ (UNCORRECTED)	I_θ (OBSERVED)
40°	3.36	3.47	3.41
60°	1.79	1.89	1.80
80°	1.14	1.29	1.13
90°	1.06	1.22	1.02
100°	1.02	1.22	1.00
120°	1.13	1.42	1.13
140°	1.29	1.72	1.37
155°	1.47	1.98	1.55

for the case of short wave-lengths as well as for long wave-lengths. Of course, this is nothing new and similar cases are well known in modern spectral theories.

It is interesting to note that the correction factor in consideration is directly taken over from the theory of the scattering of x-rays by free electrons. This is not quite in accordance with the current view[10] that the incoherent or modified scattered radiation should consist of several components having frequencies corresponding to the transitions from the initial state of the scattering atom to the final states of which the energies may lie in either the discrete or the continuous range of possible energy values. In view of the fact, however, that experiments performed by Coster,[11] Ehrenberg,[12] Kast,[13] Bearden[14] and Gingrich[15] fail to detect any fine structure in the scattered x-radiation, the introduction of this factor in the way as done by Compton seems to be justified.

It may be mentioned that, if experimental difficulties could be overcome, absolute measurements of the scattered intensity at large angles of scattering for comparatively short wave-length should give a crucial test of the present theory.

In conclusion, the writer wishes to express his appreciation of the inspiration and unfailing interest of his former teacher, Prof. A. H. Compton of University of Chicago, U. S. A.

[1] Y. H. Woo, *Proc. Nat. Acad. Sci.*, **17**, 467, 1931.

[2] C. S. Barrett, *Ibid.*, **14**, 20, 1928.

[3] C. S. Barrett, *Phys. Rev.*, **32**, 22, 1928.

[4] A. H. Compton, *Ibid.*, **35**, 925, 1930.

[5] These values of s are calculated according to Rasetti's measurements of the moment of inertia, cf. *Phys. Rev.*, **34**, 367, 1930.

[6] G. Breit, *Ibid.*, **27**, 243, 1926.

[7] P. A. M. Dirac, *Proc. Roy. Soc., AIII,* 405, 1926.
[8] Mark and Schocken, *Naturwiss.*, **15,** 139, 1927.
[9] Cf. Y. H. Woo, *Proc. Nat. Acad. Sci.*, **16,** 814, 1930; *Nature,* Oct. 4, 1930.
[10] Cf. G. Wentzel, *Zeits. Phys.*, **43,** 1 and 779, 1927; and I. Waller, *Ibid.*, **51,** 213, 1928.
[11] D. Coster, *Nature,* **123,** 642, 1929.
[12] W. Ehrenberg, *Zeits. Phys.*, **53,** 234, 1929.
[13] W. Kast, *Ibid.*, **58,** 519, 1929.
[14] J. A. Bearden, *Phys. Rev.*, **35,** 1426, 1930.
[15] N. S. Gingrich, *Ibid.*, **36,** 1050, 1930.

NOTE ON SCATTERING BY DIATOMIC GASES

BY E. O. WOLLAN

RYERSON PHYSICAL LABORATORY, UNIVERSITY OF CHICAGO

Communicated July 9, 1931

On page 467 of this issue Woo has made a correction of the general relation of Debye[1] for determining the intensity of scattering of x-rays by a gas molecule. Debye's relation is derived from classical theory and no account is taken of the modified radiation. Assuming that the modified radiation is incoherent, Woo has used the relation given by Compton[2] to correct Debye's equation. For a diatomic gas of two like atoms he gets for the intensity of scattering

$$S = \frac{I_\varphi}{ZI_e} = \frac{1}{Z}\left\{\left(1 + \frac{\sin x}{x}\right)F^2 + \frac{Z - \frac{F^2}{Z}}{(1 + \gamma\ vers\ \varphi)^3}\right\}. \qquad (1)$$

Where I_e is the independent scattering by a single electron, $x = \frac{4\pi l \sin\frac{\varphi}{2}}{\lambda}$, l being the distance between the atomic centers, F is the atomic structure factor and $\gamma = \frac{h}{mc\lambda}$.

To make calculations of the scattering by a gas molecule from the above relation it is necessary to know the structure factors for the individual atoms. A simple method of getting approximate values of F has been developed independently by Thomas[3] and Fermi.[4] Woo has calculated the structure factors by this method and substituted them into equation (2) to get the intensity of scattering from H_2, O_2 and N_2. He then compares the values with the experimental data of Barrett for these gases.

It was thought worth while to compare these calculations with the more

recent scattering data of the writer[5] which was not available to Woo at the time he wrote his papers.

Let us consider first the above-mentioned method for determining the structure factor. Thomas and Fermi have made a table from which the

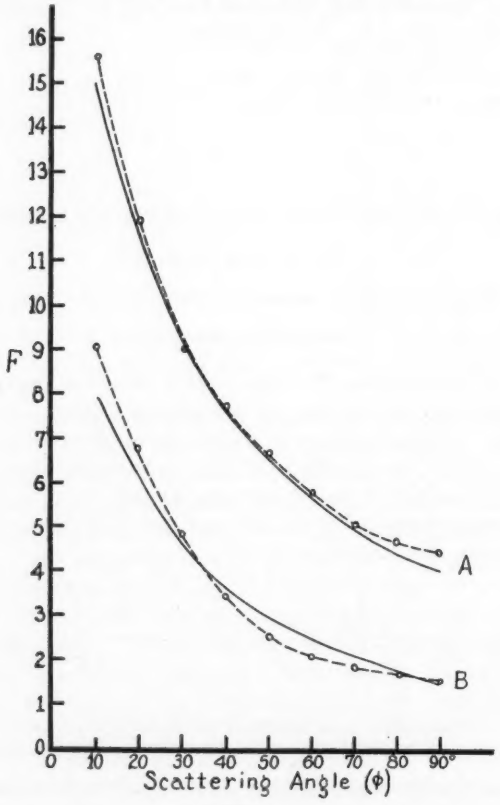


FIGURE 1

Broken line from experiment; solid line theoretical.
Curve A, argon; Curve B, neon.

structure factors can be obtained for an element of any atomic number. It takes no account of the concentration of electrons in various shells but handles the electrons as a degenerate gas surrounding the nucleus of the atom. Wherever there is a distinct separation between the electrons in the K and L shells, the method will probably not give very accurate results. In figure 1, curve B shows the structure factor curve for neon gas as determined experimentally by the writer and also the corresponding curve

calculated by the above method. The agreement is not especially good, which is probably due to the somewhat distinct separation of the K from the L electrons in the atom, as shown elsewhere.[6] In curve A a similar comparison is made for Argon, and here it is seen that the agreement is

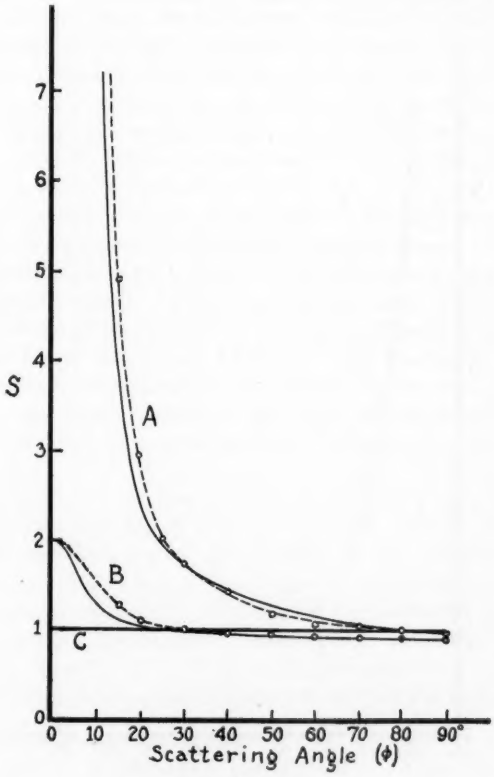


FIGURE 2

Broken lines from experiment; solid lines theoretical. Curve A, oxygen; Curve B, hydrogen. C. Thomson theory for independent scattering by single election.

quite satisfactory except at 90°. In this case the curve has not been extended to large enough values of φ to get much effect of the K electrons. In general, it seems that this method gives satisfactory values for the structure factor of the heavier elements.

Another difficulty which is present in this type of calculation of the scattering by a molecule is the assumption that the atoms maintain their

spherical symmetry. This may be a good approximation for heavy atoms where the distance of separation is large compared to the atomic radii, but will probably be only a rough approximation for N_2 and O_2 and will be a poor approximation for H_2.

However, if equation (1) gives a good representation of the interference

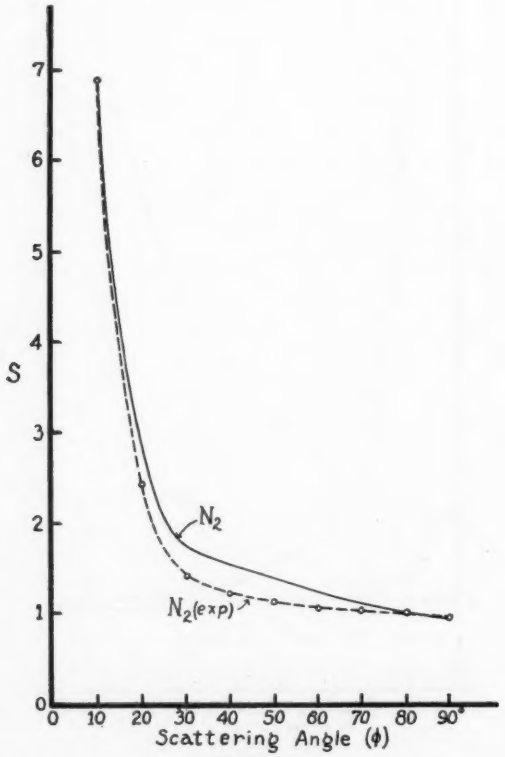


FIGURE 3

Nitrogen. Broken line from experiment; solid line theoretical.

between the atoms within the molecule one would expect fair agreement with experiment even if the above assumptions are not entirely justified.

Figures 2 and 3 show a comparison of the scattering by H_2, O_2 and N_2 as calculated by equation (1), with the experimental data of the writer.

In the experimental curve for H_2 the interference sets in at a larger angle than shown by the calculated scattering curve. In view of the fact that the electrons in the H_2 molecule are more intimately associated than is

assumed in a dumb-bell type of molecule this is precisely what would be expected.

The experimental curve for O_2 is in fair agreement with the theoretical curve and it is interesting to note that for both N_2 and O_2 the calculated intensity at 90° is identical with that determined by experiment.

The experimental curve for N_2 diverges considerably from the calculated curve between 30° and 70°. However, a noticeable bend in both curves at about 30° probably represents the point at which molecular interference sets in, and it is likely that the difference between the two curves at larger angles is due to an inaccuracy in the calculated value of F. This is borne out by the fact that the experimental and calculated values of F for neon differ in the same manner in this region.

In figure 2 the heavy horizontal line at $S = 1$ corresponds to the scattering by a single free electron on Thomson's classical theory. The calculated and experimental curve for hydrogen differs from this at large angles where there is no interference, by just the amount given by the Breit-Dirac theory. In view of this agreement and the agreement of the uncorrected equation of Debye for heavier elements where the modified radiation is negligible and the calculated structure factors are in better agreement with the theory, one seems justified in believing that equation (1) should be valid wherever one can assume spherical symmetry for the atoms in the gas molecule.

It is of interest then to reverse the line of reasoning and use equation (1) to give the structure factor and the resultant electron distribution for atoms from measurements of the intensity of scattering from molecules of the O_2 type. This assumes, however, that the distance between atoms is known from other sources. If equation (1) is reliable, as it appears to be, it extends the field of study of atomic structure factors from measurements of the intensity of x-ray scattering by gases to a range including not only the monatomic gases but also the simpler gas molecules.

[1] P. Debye, *Zeit. Phys.*, **31**, 419 (1930).
[2] A. H. Compton, *Phys. Rev.*, **35**, 925 (1930).
[3] L. H. Thomas, *Proc. Camb. Phil. Soc.*, **23**, 542 (1927).
[4] E. Fermi, *Zeit. Phys.*, **48**, 73 (1928).
[5] E. O. Wollan, *Phys. Rev.*, **37**, 862 (1931).
[6] E. O. Wollan, *Ibid*, July 1, 1931.

THE SOLID SOLUTION THEORY OF DYEING*

BY WILDER D. BANCROFT AND JOHN W. ACKERMAN

BAKER CHEMICAL LABORATORY, CORNELL UNIVERSITY

Communicated July 9, 1931

In 1890 Otto N. Witt[1] advanced the theory that the dye dissolves in the fabric or the mordant, forming a solid solution. The chief argument in favor of this view was that the color of the dye on the fiber is that of the dissolved dye and not of the solid dye. Magenta, for instance, dyes a red and not a metallic green. Silk dyed with rhodamine fluoresces, while solid rhodamine does not, when obtained on a glass plate by evaporation of an alcoholic solution.

We know now that dyeing is an adsorption phenomenon; but no one has met specifically the two points raised by Witt. The answer has recently been given implicitly in a paper by Weiser and Porter,[2] who showed that alumina takes up alizarin in the form of the alizarin anion, because one gets the same red lake whether the lake is treated with acid or alkali, whereas solid sodium alizarate is purple and solid alizarinic acid is yellow.

It seems probable that we might have non-ionizing adsorbents as well as ionizing adsorbents and this has proved to be the case. We[3] first confirmed the results of Weiser and Porter with alizarin and then tried the effect of other mordants. With stannic oxide alizarinic acid dissolved in alcohol gives a yellow-orange lake while an alkaline aqueous solution of sodium alizarate gives a purple lake. Consequently stannic oxide adsorbs chiefly undissociated alizarinic acid and undissociated sodium alizarate. Incidentally, sodium alizarate, though a polar compound, most be sufficiently undissociated to give a purple color. An x-ray study of this compound would be rather worth while.

With hydrous zinc oxide there is a slight adsorption of yellow alizarinic acid and purple sodium alizarate. A variation in the degree of acidity with stannic oxide and zinc oxide will give all variations of color between yellow and purple, while the presence of alumina introduces a red. This accounts for the fact of alizarin being what is called a polygenetic color,[4] because the color varies with the nature of the mordant. Other examples of this class are said to be cochineal and logwood; but we have not yet studied these dyes.

Rhodamine 6G does not fluoresce in the solid state but shows a green-yellow fluorescence in aqueous solution. It dyes silk, wool and cotton a yellow-red; but the dyed silk shows a strong yellow fluorescence, the dyed wool practically no fluorescence and the dyed cotton no fluorescence. Consequently we deduce that rhodamine 6G is distinctly ionized on silk, only very slightly on wool and not at all on cotton.

Some experiments were also made with lakes of rhodamine 6G. With the stannic oxide lake there was no fluorescence; with the zinc oxide and alumina lakes there was a slight yellow fluorescence; and with the silica lake there was a distinct yellow fluorescence. Consequently silica is an ionizing adsorbent for rhodamine 6G, while stannic oxide is a non-ionizing adsorbent. The zinc oxide and the alumina lakes ionize rhodamine 6G to some extent but not so much as silica does.

Experiments with rhodamine B gave results that were similar in type; but we are not able at present to account for the colors of the fluorescence in the several cases. Rhodamine B does not fluoresce in the solid state but shows a brownish-yellow fluorescence in aqueous solution. On silk the color is blue-red with a strong red fluoresecnce, on wool a blue-red with a very slight fluorescence and on cotton a blue-pink with no fluorescence. The stannic oxide lake is purple and does not fluoresce. The zinc oxide and the alumina lakes are pink and show a slight yellow fluorescence. The silica lake is pink and shows a distinct yellow fluorescence.

The general conclusions of this paper are as follows: 1. An alcoholic solution of alizarinic acid added to alumina gives a pink lake, showing the adsorption of the red alizarin anion.

2. With alumina sodium alizarate gives a red lake, which does not change color when acidified. The color is due, as Weiser has said, to alumina being an ionizing adsorbent for alizarin.

3. The hydrous oxide of tin adsorbs undissociated alizarinic acid to give an orange lake. The yellow is due to undissociated alizarinic acid with probably just enough alizarin ion to account for the orange color.

4. The hydrous oxide of tin adsorbs undissociated sodium alizarate to give a purple lake. There may be some alizarin ion present.

5. Hydrous zinc oxide adsorbs undissociated sodium alizarate to form a purple lake. There is no way at present to tell whether there is some alizarin ion present or not.

6. Since sodium alizarate, a polar compound, contains an unknown amount of non-ionized salt, a systematic x-ray study of this compound is much to be desired.

7. By changing the mordant and the acidity the color of alizarin lakes can be made to vary from purple to red to yellow with a colorless mordant. This explains why alizarin is called a polygenetic dye.

8. Silk dyed with rhodamine 6G is yellow-red with a strong yellow fluorescence; on wool and cotton the color is yellow-red with little or no fluorescence.

9. Silk dyed with rhodamine B is blue-red with a strong red fluorescence: wool is blue-red with a very slight fluorescence; and cotton is blue-pink with no fluorescence.

10. The tin lakes of rhodamine B and rhodamine 6G show no fluores-

cence; the alumina and zinc lakes show a slight yellow fluorescence; and the silica lakes exhibit a distinct yellow fluorescence.

11. Alumina is a strongly ionizing adsorbent with alizarin and a weakly ionizing adsorbent with the rhodamines.

12. Witt's theory of solid solutions in dyeing loses its only support as soon as one postulates that textiles and mordants may act as ionizing adsorbents.

* This work is done under the programme now being carried out at Cornell University and supported in part by a grant from the Heckscher Foundation for the Advancement of Research established by August Heckscher at Cornell University.

[1] O. N. Witt, *Chem. Centralbl.*, [4] **2**, II, 1039 (1890).

[2] Weiser and Porter, *J. Phys. Chem.*, **31**, 1824 (1927).

[3] Bancroft and Ackerman, *Ibid.*, **35**, 2568 (1931).

[4] Hummel, *The Dyeing of Textile Fabrics*, p. 147, 1885.

---

*REVERSIBLE COAGULATION IN LIVING TISSUE**

BY WILDER D. BANCROFT AND JOHN E. RUTZLER, JR.**

BAKER CHEMICAL LABORATORY, CORNELL UNIVERSITY

Communicated July 9, 1931

According to Claude Bernard[1] "the action of anesthetics is very general. They react not only with animals but also with plants." Bose[2] has found that plants anesthetized with ether go through a preliminary period of increased irritability or stimulation. A dilute alcohol solution acted as a stimulant, whereas a concentrated alcohol solution acted as a depressant. Nichols[3] applied chloroform to the wall of an internode of *Nitella* by means of a capillary pipette. An area of non-motile protoplasm was produced at the point of application of the anesthetic. This area of non-motile protoplasm resumed its movements later, so the process appears to be a reversible one. Nichols considers the effect to be due to gelation, thus providing us with a case of reversible coagulation in plants accompanied by anesthesia of the part affected.

Wallace[4] has recently studied the effect of temperature upon the sensitive plant by measuring the angle of movement of the branches. The sensitivity of *Mimosa* is lost at 60°, which is about the coagulation temperature of albumin. The sensitivity also becomes zero at 12.5°, a temperature at which there can be no ice crystals. The temperature of maximum activity was about 40°. Frogs can be made completely insensitive to external influences by putting them in water at 0°, and they are also subject to heat narcosis in very warm water. There is a striking parallelism between plants and animals in their responses to heat and cold coagulation. The

experiments of Wallace are an additional confirmation of the view that the reversible coagulation of proteins is the important factor in anesthesis.

It seems a natural corollary of our point of view that the behavior of the sensitive plant, *Mimosa pudica*, was due to the proteins being in the partially coagulated stage. Since nobody knows what the proteins are, it was impossible to tell from a pH determination how near the isoelectric point they were. An indirect line of attack was therefore adopted, based on the assumption that a sodium thiocyanate solution should counteract the sensitivity to some extent. Professor Knudsen of the New York State College of Agriculture was good enough to grow a number of sensitive plants for us and to help us considerably with the experiments.

When one of the plants is sprayed with a five per cent sodium thiocyanate solution and the roots are later watered with the same solution, it becomes relatively insensitive to mechanical stimulation; even comparatively powerful blows elicited only a faint folding response on the part of the leaves and the petioles dropped only slightly. The order of sensitivity of the plants as nearly as could be determined after spraying and watering was: sodium thiocyanate < sodium salicylate < control > sodium citrate > sodium amytal. This is what one would expect if one postulates that the sensitive plant is nearly at the maximum sensitivity in its normal state. Both dispersion and coagulation will then produce a decrease in irritability. The unexpected thing is that sodium amytal is apparently nothing like as powerful an anesthetic as ether or chloroform.

In order to show the postulated antagonism between ether and sodium thiocyanate, 5 cc. of ether were poured on the soil in which a potted sensitive plant was growing, care being taken that the ether did not come in direct contact with the plant. One leaf closed up and the petioles fell immediately. Complete anesthesia ensued in four minutes, the leaves and petioles both being so affected that the plant was completely folded up. The sensitivity of the pulvini began to return several times as evidenced by an extremely slight rise of the petioles; but each time they fell again. After one hour the pulvini were slightly sensitive and the petioles had risen a few millimeters; the leaves had not opened at all.

Along with this experiment a potted plant, that had been sprayed twice with a five per cent solution of sodium thiocyanate and that had stood for about thirty minutes after the second spraying, was treated by pouring 5 cc. of ether on the soil, being careful not to touch the plant. After six minutes the anesthetic had a slight effect; but the leaves were not acted on at all so far as one could see, for they did not close. In the case of the control (or ether) plant the petioles dropped until they were supported by the edge of the pot, in the case of the plant treated with sodium thiocyanate the petioles dropped about one quarter as far and were fully erect again seventeen minutes after they fell.

A third plant was anesthetized by pouring 5 cc. of ether on the soil. It took five and one-half minutes for the ether to take effect, at which time the petioles fell and the leaves closed. After forty minutes the pulvini were slightly sensitive as evidenced by a few millimeters rise of the petioles and by their dropping again when touched. At the same time that this third plant was treated, 5 cc. of ether were poured upon the soil of a fourth plant. The leaves closed and the petioles fell in two and one-half minutes. One minute and a half later the plant was sprayed with a five per cent solution of sodium thiocyanate and in eighteen minutes the leaves were partly open, the petioles were partly raised and the pulvini were quite sensitive.

The general results of this paper are as follows:

1. In plants as in animals the effect of anesthetization is a reversible coagulation of the proteins.
2. Plants pass through a preliminary stage of excitation when going from the normal state to the state of anesthesia.
3. In *Mimosa pudica* some of the proteins are apparently very close to the state of maximum excitation.
4. The sensitivity of the sensitive plant can be decreased by the addition of sodium thiocyanate and sodium salicylate which are peptizing agents, and by the addition of sodium citrate and amytal which are coagulating agents.
5. Contrary to what has been found with animals, sodium amytal is apparently not an efficient anesthetizing agent with the sensitive plant.
6. Sodium thiocyanate minimizes the anesthetic effect on the sensitive plant, wherether added before or after the administration of the ether.
7. Claude Bernard's theory of anesthesia has proved applicable to plants as well as to animals.

* This work is done under the programme now being carried out at Cornell University and supported in part by a grant from the Heckscher Foundation for the Advancement of Research established by August Heckscher at Cornell University.

** Eli Lilly Fellow.

[1] Claude Bernard, *Leçons sur l'anesthésiques et sur l'asphyxie* (1875).

[2] Bose, *Researches on Irritability of Plants*, 334 (1913).

[3] Nichols, *Bull. Torrey Bot. Club*, **57**, 152 (1930).

[4] Wallace, *Am. J. Bot.*, **18**, 293 (1931).

THE ORDER OF THE GENES C, Sh AND Wx IN ZEA MAYS WITH REFERENCE TO A CYTOLOGICALLY KNOWN POINT IN THE CHROMOSOME

BY BARBARA MCCLINTOCK

BOTANY DEPARTMENT, CORNELL UNIVERSITY

Communicated July 7, 1931

The monoploid complement in *Zea mays* is composed of ten morphologically distinguishable chromosomes.[1] By means of trisomic inheritance it has been possible to associate certain linkage groups with particular members of the complement.[2] Thus the linkage group involving the genes *C* (colored aleurone), *sh* (shrunken endosperm) and *wx* (waxy starch) has been associated with the next to the smallest chromosome (chromosome 9) of the monoploid complement. It is the purpose of this paper to present evidence indicating the serial order of these genes with reference to a known point in the chromosome.

In a previous report[3] it was shown that in a certain strain of maize a segmental interchange had occurred between chromosomes 8 and 9. The diagram in figure 1 is similar to the one accompanying that report. By means of the close synaptic attraction of the homologous parts of chromosomes in early meiotic prophase it has been possible to determine the region in each chromosome at which the interchange occurred, and to recognize this region in the sporocytes in plants with the chromosome complex shown in *c*, figure 1. The conspicuous knob terminating the short arm of chromosome 9 (*n* in the diagram) is found only in certain maize cultures. However, in these cultures it is a constant feature of this chromosome. Evidence presented in this paper indicates that the loci of the genes *wx*, *sh* and *c* lie in the region between the interchange point and the knob, *wx* being nearest the interchange and *c* farthest from it.

Plants which are heterozygous for the interchange possess the chromosomes *nNIi* (see diagram) and show approximately 50% sterility in pollen and eggs.[4] At diakinesis the interchange complex appears as a ring of four chromosomes (*d*, Fig. 1). The four members of this ring distribute 2-by-2 at anaphase. Since in half of the sporocytes any two adjacent chromosomes in the ring go to the same pole, forming sterile combinations, and in the other half of the sporocytes and the adjacent members go to opposite poles, forming fertile combinations, six types of spores are formed, only two of which are viable: those containing *n* with *N* and those containing *I* with *i*. Occasionally the distribution at anaphase is such that 11 chromosomes, including 3 members of the ring, go to one pole and 9 to the other. In this manner viable 11-chromosome gametes are formed. Such gametes could possess the following chromosomes: (1) *iIN*, (2) *iIn*, (3) *nNI*, (4) *nNi*.

Through fusion of these 11-chromosome gametes with normal gametes four types of 21-chromosome plants can be obtained. The types of chromosome configurations seen at diakinesis in the resulting 21-chromosome individuals enable one to determine which of the four types of 11-chromosome gamete has functioned in any particular case. Experience has shown that fusion of a normal gamete with one of type (1) or type (2) will produce an individual which is approximately 25–30% sterile;[4] where-

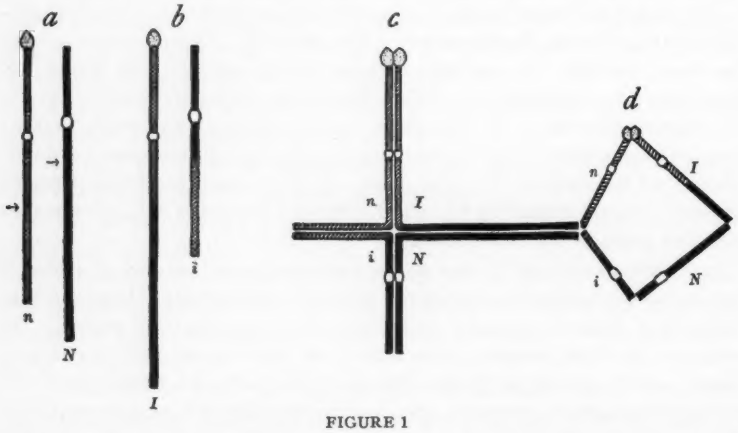


FIGURE 1

a.—Diagram of the two normal chromosomes which were involved in the segmental interchange. The clear portions in the chromosomes represent the spindle fibre attachment regions. The smaller chromosome terminated in an enlarged, deeply staining knob. The arrows indicate the places in the chromosome at which the interchange occurred to produce the situation shown in *b*. *b*.—The two chromosomes produced as the result of the segmental interchange. *c*.—The type of synaptic complex in midprophase of meiosis obtained by combining a normal chromosome complement with an interchange complement through crossing. *N*, larger normal chromosome; *n*, smaller normal chromosome; *I*, larger interchange chromosome; *i*, smaller interchange chromosome. *d*.—The ring formed at diakinesis by the opening out of chromosomes from the complex shown in *c*. (Chromosomes labeled as in C.)

as, fusion of a normal gamete with a gamete of type (3) or type (4) will produce an individual which is approximately 18% sterile. In the former cases the chromosome configuration at diakinesis is frequently a closed ring of four chromosomes with the extra chromosome attached. In the latter cases the configuration is not a ring, but consists either of a "chain" of five chromosomes, a bivalent plus a trivalent or, infrequently, 2 bivalents plus a univalent. The "chain" is composed of two bivalent chromosomes, bivalent 8 (*N*) and bivalent 9 (*n*), held together by the extra chromosome (*I* or *i*) which is synapsed at one end with 8 (*N*) and at the other end with 9

(*n*) (Fig. 2). Opening out of the members of the bivalents in late prophase produces a chain-like structure.

Plant 270_{11} belonged to the low sterile class. This plant was suspected of having the formula *nnNNI* since it was noted that in those sporocytes which showed a trivalent instead of a chain of five chromosomes, the trivalent always consisted of one long and two small chromosomes. Since no knob was present at the end of either the *n* or the *I* chromosome it was impossible to be certain from such observations. Pollen counts from several anthers indicated that the plant was approximately 18% sterile. If this plant had the formula *nnNNI*, four types of functional gametes would be produced: (1) *nN*, (2) *nNN*, (3) *nnN*, (4) *nNI*. Since 11-chromosome pollen grains seldom function in competition with grains containing only the monoploid complement, selfing such a plant would produce four types of individuals. All the 20-chromosome individuals would be ordinary normals. Of the 21-chromosome individuals there would be three kinds. The functioning of gametes of types (2) and (3) would produce individuals trisomic for chromosomes 8 (*N*) and 9 (*n*), respectively. The diakinetic configuration shown by such individuals would be nine bivalents plus a trivalent or ten bivalents plus a univalent. No sterility above that found in normal diploid individuals (approximately 2–5%) would be observed in pollen or eggs. The functioning of type (4) would produce an individual of the same chromosome constitution as the parent (*nnNNI*). It would show a chain of five chromosomes at diakinesis and exhibit 18% sterility in pollen and eggs.

Twenty-two individuals resulting from selfing plant 270_{11} were grown and examined cytologically. Eight of these were 20-chromosome individuals. They were ordinary diploids, showed no sterility and gave disomic ratios for *c*, *sh* and *wx*. Twelve of the thirteen 21-chromosome individuals must have come from gametes of type (4) since they showed 18% sterility in pollen and eggs. Cytological observations indicated that the remaining 21-chromosome individual was trisomic for chromosome 8 (*N*) and therefore must have involved a gamete of type (2). It was non-sterile and gave disomic inheritance for the genes *c*, *sh* and *wx*.

On the basis of the appearance of the chromosomes at diakinesis the assumption has been made that the formula for the parent plant and all its partially-sterile offspring was *nnNNI*. To test this assumption several of the partially sterile offspring were crossed to ordinary normal plants whose number 9 (*n*) chromosomes terminated in conspicuous knobs. Chromosome counts were obtained from nineteen individuals resulting from such a cross. There were nine 20-chromosome individuals, all normals. Among the ten 21-chromosome individuals seven were partially sterile like the female parent and three were non-sterile. Two of the three non-sterile 21-chromosome individuals were examined and were found to be

trisomic for chromosome 8 (*N*). Among the partially sterile plants two were examined for chromosome configurations at diakinesis. In each of them a chain of five chromosomes was observed. If the extra chromosome in these 21-chromosome plants were the long interchange chromosome (*I*), the knob should be found only in two positions.* These positions are illustrated in figure 2.

The conspicuous knob allows the configuration to be easily interpreted. In every sporocyte in which a chain of five chromosomes appeared the knob was found in one or the other of the positions illustrated. In some sporocytes a trivalent plus a bivalent appeared instead of a chain of five. The presence of the knob on one of the *n* chromosomes makes it possible to state that such a trivalent is composed either of *nnI* or *NNI*. This evi-

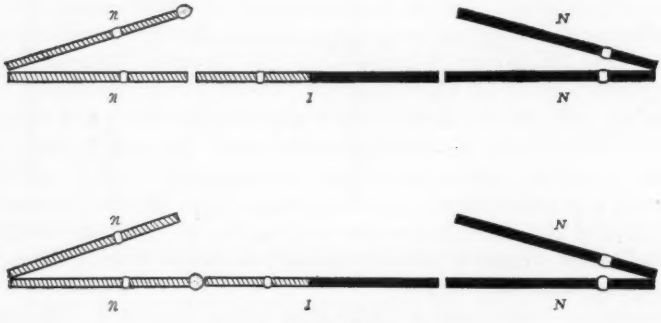


FIGURE 2

dence makes it certain that the extra chromosome in the complement of these sterile plants is the long interchange chromosome (*I*).

The configuration which would be seen if the short interchanged chromosome (*i*) were present would be a chain of five chromosomes, also. However, the long arm of chromosome 9 would be associated with the interchanged chromosome and so the knob would be where the two *n* chromosomes join. No such case was seen.

Chromosome *I* possesses nearly two-thirds of that chromosome (number 9) which carries the genes of the *c-sh-wx* linkage group. It is highly probable, therefore, that the loci of some of the known genes of this linkage group lie in this portion of the chromosome. Crossover data have shown that the locus of the gene *sh* lies between that of *c* and *wx*. The map distance between *c* and *sh* is 3.3, that between *sh* and *wx*, 20.3.[5] Unpublished data which Dr. C. R. Burnham has generously allowed me to use have shown that the percentage of crossing-over between *wx* and the interchange point is approximately 13 per cent, between *sh* and the interchange

approximately 32 per cent, between *c* and the interchange, approximately 33 per cent. If it could be shown that the long interchange chromosome (*I*) carries the genes *wx*, *sh* and *c* or their allelomorphs, their order with reference to the interchange point could be established. The following evidence is presented in support of this claim.

TABLE 1

RATIOS OBTAINED FROM SELFING PLANT 270_{11}

COLORED				COLORLESS			
Sh		*sh*		*Sh*		*sh*	
Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*
11	8	1	0	25	5	9	2

Plant 270_{11} was heterozygous for the genes *c*, *sh* and *wx* (table 1). Among the 21-chromosome, partially sterile individuals which came from the colored, non-shrunken, non-waxy kernels of 270_{11}, three were utilized for genetic investigations. These plants were numbered $388A_1$, $388D_1$ and $388D_2$. All had the formula *nnNNI*.

The pollen of plant $388A_1$ was examined for the waxy ratio and sterility (table 2). The waxy ratio in the pollen of this plant was approximately 15 *Wx*:1 *wx*. This is in contrast to the 1 *Wx*:1 *wx* ratio found among diploid individuals or individuals which are trisomic for some chromosome other than that which carries the genes of the *c-sh-wx* linkage group. In most anthers the sterility approximates 18%. For reasons as yet unknown, certain anthers show a considerably lower or higher percentage of sterility. When this pollen was used on a plant homozygous for *c*, *sh* and *wx* the following ratio was obtained:

296_9 ♀ × $388A_1$ ♂

C				*c*			
Sh		*sh*		*Sh*		*sh*	
Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*
27	22	3	0	0	0	172	5**

When crossed to a plant heterozygous for *c* but homozygous for *sh* and *wx* the following resulted:

$389C_1$ ♀ × $388A_1$ ♂

C				*c*			
Sh		*sh*		*Sh*		*sh*	
Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*
41	26	124	6	0	0	119	5**

In examining these data it should be remembered (1) that the functional gametes formed by plant $388A_1$ contain *nN*, *NNn*, *nnN* or *nNI* in varying

proportions according to the mode of distribution at meiosis (page 487), (2) that the spores with *nI* or *NI* abort and are the primary cause of the observed sterility and (3) that the 11-chromosome pollen grains do not function in competition with those containing the monoploid set. The ratio for waxy in the pollen and the ratios and linkage relations of the genes *c*, *sh* and *wx* in the backcross are distinctly abnormal. The ratios are explainable on the assumption that the interchange chromosome (*I*) has that portion of chromosome 9 which involves the loci of the genes *c*, *sh* and *wx*. Accordingly, it can be stated that plant $388A_1$ possessed two *n* chromosomes each with the genes *c-sh-Wx* and an interchange chromosome (*I*) with the genes *C-Sh-wx*.

The genic constitution of the two *n* chromosomes in plant $388D_1$ were *c-sh-wx* and *C-sh-Wx*, respectively; chromosome *I* contained the genes *C-Sh-wx*. This is clear from the following cross to the triple recessive:

296_{13} ♀ × $388D_1$ ♂

C				*c*			
Sh		*sh*		*Sh*		*sh*	
Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*
11	21	72	11	0	1	8	79

That one *n* chromosome of plant $388D_2$ contained the genes *c-sh-Wx*, the other *n* chromosome *C-Sh-Wx* and the long interchange chromosome (*I*) the genes *C-Sh-wx* can be seen from the following crosses:

$388D_2$ ♀ × 390_3 (homozygous *c-sh-wx*) ♂

C				*c*			
Sh		*sh*		*Sh*		*sh*	
Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*
108	3	3	0	1	0	38	0

296_{22}(homozygous *c-sh-wx*) ♀ × $388D_2$ ♂

C				*c*			
Sh		*sh*		*Sh*		*sh*	
Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*
194	34	2	0	5	0	130	4**

296_{21}(homozygous *c-sh-wx*) ♀ × $388D_2$ ♂

C				*c*			
Sh		*sh*		*Sh*		*sh*	
Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*
197	18	3	2	3	0	150	5**

The above results point to the fact that the loci for the genes *c*, *sh* and *wx* lie in that part of the long interchange chromosome (*I*) contributed by chromosome 9. That the genes lie in the order *wx-sh-c* beginning at the interchange point is indicated by the fact that *wx* shows 13%, *sh* 32% and *c* 33% crossing over from the interchange point. It is not known on which side of the spindle-fibre insertion region these genes lie. It is possible, however, that they may like in the short arm, between the insertion region and the knob, since evidence presented by Stadler[6] suggests that the gene *wx* lies closer to the spindle-fibre insertion region than the genes *sh* or *c*.

TABLE 2

POLLEN COUNTS FROM INDIVIDUAL ANTHERS OF PLANT 388 A_1

Wx	*wx*	ABORTIVE	PER CENT ABORTIVE	PER CENT *wx*
991	59	129	10.9	5.6
2029	142	937	30.1	6.5
632	48	145	17.5	7.0
701	55	164	18.0	7.2
1035	80	240	17.7	7.1
1415	69	347	18.9	4.7

The author wishes to state that no attempt has been made to discuss in detail the data herein presented. It was desired to present briefly the evidence at this time, since it lends valuable support to the argument in the paper which follows.

* The position of the knob resulting from crossing-over will be described in a later paper.

** Among the plants of the *c-sh-wx* cultures four or five silks were exserted before the ear shoot appeared above the leaf sheath and was bagged. It is possible that the high number of *c-sh-wx* kernels mostly represent self-pollinations rather than double crossovers or crossovers involving all three chromosomes. Similar crosses are being repeated.

[1] McClintock, B., *Sci.*, **69**, 629 (1929).

[2] McClintock, K. B., and Henry E. Hill, *Genetics*, **16**, 175–190 (1931).

[3] McClintock, B., *Proc. Nat. Acad. Sci.*, **16**, 791–796 (1930).

[4] Burnham, C. R., *Ibid.*, **16**, 269–277 (1930).

[5] Mimeographed pamphlet on linkage in maize. Cornell University.

[6] Stadler, L. J., *Sci. Agr.*, **11**, 557–572 (1931).

A CORRELATION OF CYTOLOGICAL AND GENETICAL CROSSING-OVER IN ZEA MAYS

By Harriet B. Creighton and Barbara McClintock

Botany Department, Cornell University

Communicated July 7, 1931

A requirement for the genetical study of crossing-over is the heterozygous condition of two allelomorphic factors in the same linkage group. The analysis of the behavior of homologous or partially homologous chromosomes, which are morphologically distinguishable at two points, should show evidence of cytological crossing-over. It is the aim of the present paper to show that cytological crossing-over occurs and that it is accompanied by genetical crossing-over.

In a certain strain of maize the second-smallest chromosome (chromosome 9) possesses a conspicuous knob at the end of the short arm. Its distribution through successive generations is similar to that of a gene. If a plant possessing knobs at the ends of both of its 2nd-smallest chromosomes is crossed to a plant with no knobs, cytological observations show that in the resulting F_1 individuals only one member of the homologous pair possesses a knob. When such an individual is back-crossed to one having no knob on either chromosome, half of the offspring are heterozygous for the knob and half possess no knob at all. The knob, therefore, is a constant feature of the chromosome possessing it. When present on one chromosome and not on its homologue, the knob renders the chromosome pair visibly heteromorphic.

In a previous report[1] it was shown that in a certain strain of maize an interchange had taken place between chromosome 8 and 9. The interchanged pieces were unequal in size; the long arm of chromosome 9 was increased in relative length, whereas the long arm of chromosome 8 was correspondingly shortened. When a gamete possessing these two interchanged chromosomes meets a gamete containing a normal chromosome set, meiosis in the resulting individual is characterized by a side-by-side synapsis of homologous parts (see diagram, figure 1 of preceding paper). Therefore, it should be possible to have crossing-over between the knob and the interchange point.

In the previous report it was also shown that in such an individual the only functioning gametes are those which possess either the two normal chromosomes (N, n) or the two interchanged chromosome (I, i), i.e., the full genom in one or the other arrangement. The functional gametes therefore possess either the shorter, normal, knobbed chromosome (n) or the longer, interchanged, knobbed chromosome (I). Hence, when such a plant is crossed to a plant possessing the normal chromosome complement,

the presence of the normal chromosome in functioning gametes of the former will be indicated by the appearance of ten bivalents in the prophase of meiosis of the resulting individuals. The presence of the interchanged

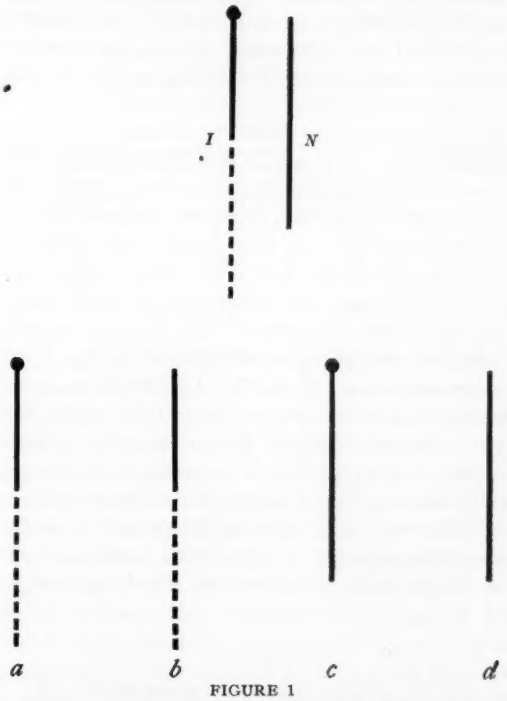


FIGURE 1

Above—Diagram of the chromosomes in which crossing-over was studied. (Labeled as in figure 1, preceding paper.)

Below—Diagram of chromosome types found in gametes of a plant with the constitution shown above.

a—Knobbed, interchanged chromosome.
b—Knobless, interchanged chromosome.
c—Knobbed, normal chromosome.
d—Knobless, normal chromosome.
a and *d* are non-crossover types.
b and *c* are crossover types.

chromosome in other gametes will be indicated in other F_1 individuals by the appearance of eight bivalents plus a ring of four chromosomes in the late prophase of meiosis.

If a gamete possessing a normal chromosome number 9 with no knob,

meets a gamete possessing an interchanged chromosome with a knob, it is clear that these two chromosomes which synapse along their homologous parts during prophase of meiosis in the resulting individual are visibly different at each of their two ends. If no crossing-over occurs, the gametes formed by such an individual will contain either the knobbed, interchanged chromosome (*a*, Fig. 1) or the normal chromosome without a knob (*d*, Fig. 1). Gametes containing either a knobbed, normal chromosome (*c*,

TABLE 1

KNOB-INTERCHANGED / KNOBLESS-NORMAL × KNOBLESS-NORMAL, CULTURE 337 AND KNOBBED-NORMAL CULTURES A125 AND 340

	PLANTS POSSESSING 2 NORMAL CHROMOSOMES		PLANTS POSSESSING AN INTERCHANGED CHROMOSOMES	
CULTURE	NON-CROSSOVERS	CROSSOVERS	NON-CROSSOVERS	CROSSOVERS
337	8	3	6	2
A125	39	31	36	23
340	5	3	5	3
Totals	52	37	47	28

Fig. 1) or a knobless, interchanged chromosome (*b*, Fig. 1) will be formed as a result of crossing-over. If such an individual is crossed to a plant possessing two normal knobless chromosomes, the resulting individuals will be of four kinds. The non-crossover gametes would give rise to individuals which show either (1) ten bivalents at prophase of meiosis and no knob on chromosome 9, indicating that a gamete with a chromosome of type *d* has functioned or (2) a ring of four chromosomes with a single conspicuous knob, indicating that a gamete of type *a* has functioned. The crossover types will be recognizable as individuals which possess either (1) ten bivalents and a single knob associated with bivalent chromosome 9 or

TABLE 2

KNOB-*C-wx* / KNOBLESS-*c-Wx* (*c-Wx*) × KNOBLESS-*c-wx*

C-wx				*C-Wx*		*c-wx*	
Knob	Knobless	Knob	Knobless	Knob	Knobless	Knob	Knobless
12	5	5	34	4	0	0	3

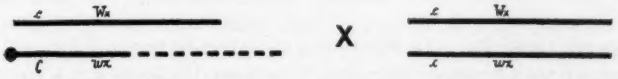
(2) a ring of four chromosomes with no knob, indicating that crossover gametes of types *c* and *b*, respectively, have functioned. The results of such a cross are given in culture 337, table 1. Similarly, if such a plant is crossed to a normal plant possessing knobs at the ends of both number 9 chromosomes and if crossing-over occurs, the resulting individuals should be of four kinds. The non-crossover types would be represented by (1) plants homozygous for the knob and possessing the interchanged chromosome and (2) plants heterozygous for the knob and possessing two normal chromosomes. The functioning of gametes which had been produced as the result of crossing-over between the knob and the interchange would give rise to (1) individuals heterozygous for the knob and possessing the

interchanged chromosome and (2) those homozygous for the knob and possessing two normal chromosomes. The results of such crosses are given in cultures Al25 and 340, table 1. Although the data are few, they are consistent. The amount of crossing-over between the knob and the interchange, as measured from these data, is approximately 39%.

In the preceding paper it was shown that the knobbed chromosome carries the genes for colored aleurone (*C*), shrunken endosperm (*sh*) and waxy endosperm (*wx*). Furthermore, it was shown that the order of these genes, beginning at the interchange point is *wx-sh-c*. It is possible, also, that these genes all lie in the short arm of the knobbed chromosome. Therefore, a linkage between the knob and these genes is to be expected.

One chromosome number 9 in a plant possessing the normal complement had a knob and carried the genes *C* and *wx*. Its homologue was knobless and carried the genes *c* and *Wx*. The non-crossover gametes should contain a knobbed-*C-wx* or a knobless-*c-Wx* chromosome. Crossing-over in region 1 (between the knob and *C*) would give rise to knobless *C-wx* and knobbed-*c-Wx* chromosomes. Crossing-over in region 2 (between *C* and *wx*) would give rise to knobbed-*C-Wx* and knobless-*c-wx* chromosomes. The results of crossing such a plant to a knobless-*c-wx* type are given in table 2. It would be expected on the basis of interference that the knob and *C* would remain together when a crossover occurred between *C* and *wx;* hence, the individuals arising from colored starchy (*C-Wx*) kernels should possess a knob, whereas those coming from colorless, waxy (*c-wx*) kernels should be knobless. Although the data are few they are convincing. It is obvious that there is a fairly close association between the knob and *C*.

To obtain a correlation between cytological and genetic crossing-over it is necessary to have a plant heteromorphic for the knob, the genes *c* and *wx* and the interchange. Plant 338 (17) possessed in one chromosome the knob, the genes *C* and *wx* and the interchanged piece of chromosome 8. The other chromosome was normal, knobless and contained the genes *c* and *Wx*. This plant was crossed to an individual possessing two normal, knobless chromosomes with the genes *c-Wx* and *c-wx*, respectively. This cross is diagrammed as follows:



The results of the cross are given in table 3. In this case all the colored kernels gave rise to individuals possessing a knob, whereas all the colorless kernels gave rise to individuals showing no knob.

The amount of crossing-over between the knob and the interchange

point is approximately 39% (Table 1), between *c* and the interchange approximately 33%, between *wx* and the interchange, 13% (preceding paper). With this information in mind it is possible to analyze the data given in table 3. The data are necessarily few since the ear contained but few kernels. The three individuals in class I are clearly non-crossover types. In class II the individuals have resulted from a crossover in region 2,

TABLE 3

$$\frac{\text{KNOB-}C\text{-}wx\text{-INTERCHANGED}}{\text{KNOBLESS-}c\text{-}Wx\text{-NORMAL}} \times \frac{\text{KNOBLESS-}c\text{-}Wx\text{-NORMAL}}{\text{KNOBLESS-}c\text{-}wx\text{-NORMAL}}$$

PLANT NUMBER	KNOBBED OR KNOBLESS	INTERCHANGED OR NORMAL	
Class I, *C-wx* kernels			
1	Knob	Interchanged	
2	Knob	Interchanged	
3	Knob	Interchanged	
Class II, *c-wx* kernels			
1	Knobless	Interchanged	
2	Knobless	Interchanged	
Class III, *C-Wx* kernels			*Pollen*
1	Knob	Normal	*WxWx*
2	Knob	Normal	....
3	....	Normal	*WxWx*
5	Knob	Normal	....
6	Knob	.....	....
7	Knob	Normal	....
8	Knob	Normal	....
Class IV, *c-Wx* kernels			
1	Knobless	Normal	*Wxwx*
2	Knobless	Normal	*Wxwx*
3	Knobless	Interchanged	*Wxwx*
4	Knobless	Normal	*Wxwx*
5	Knobless	Interchanged	*WxWx*
6	Knobless	Normal	*WxWx*
7	Knobless	Interchanged	*Wxwx*
8	Knobless	Interchanged	*WxWx*
9	Knobless	Normal	*WxWx*
10	Knobless	Normal	*WxWx*
11	Knobless	Normal	*Wxwx*
12	Knobless	Normal	*Wxwx*
13	Knobless	Normal	*WxWx*
14	Knobless	Normal	*WxWx*
15	Knobless	Normal	*Wx*—

i.e., between *c* and *wx*. In this case a crossover in region 2 has not been accompanied by a crossover in region 1 (between the knob and *C*) or region 3 (between *wx* and the interchange). All the individuals in class III had normal chromosomes. Unfortunately, pollen was obtained from only 1 of the 6 individuals examined for the presence of the knob. This one individual was clearly of the type expected to come from a gamete produced through crossing-over in region 2. Class IV is more difficult to analyze.

Plants 6, 9, 10, 13, and 14 are normal and *WxWx;* they therefore represent non-crossover types. An equal number of non-crossover types are expected among the normal *Wxwx* class. Plants 1, 2, 4, 11 and 12 may be of this type. It is possible but improbable that they have arisen through the union of a *c-Wx* gamete with a gamete resulting from a double crossover in region 2 and 3. Plants 5 and 8 are single crossovers in region 3, whereas plants 3 and 7 probably represent single crossovers in region 2 or 3.

The foregoing evidence points to the fact that cytological crossing-over occurs and is accompanied by the expected types of genetic crossing-over.

Conclusions.—Pairing chromosomes, heteromorphic in two regions, have been shown to exchange parts at the same time they exchange genes assigned to these regions.

The authors wish to express appreciation to Dr. L. W. Sharp for aid in the revision of the manuscripts of this and the preceding paper. They are indebted to Dr. C. R. Burnham for furnishing unpublished data and for some of the material studied.

[1] McClintock, B., *Proc. Nat. Acad. Sci.*, **16**, 791–796 (1930).

A REVERSE MUTATION FROM "DILUTE" TO "INTENSE" PIGMENTATION IN THE HOUSE MOUSE

BY CLYDE E. KEELER

HOWE LABORATORY OF THE HARVARD MEDICAL SCHOOL, AND THE BUSSEY INSTITUTION

Communicated June 30, 1931

For several years there have been maintained at the Bussey Institution a few cages of mice homozygous for the two closely linked recessive genes determining, respectively, short-ears and blue dilution. A litter born in one of these cages contained several short-eared non-agouti dilute mice, and, in addition, a short-eared non-agouti female showing intense pigmentation instead of the expected dilute.

Our short-eared intense stock was located in a section of the mouse room remote from the cages containing the short-eared dilute stock, making contamination from this source highly improbable. All possibility that intense pigmentation appeared as a result of contamination is completely ruled out, however, by a number of additional facts. If we suppose contamination by the short-eared intense stock to have taken place, then it could only have occurred in one of two possible ways. Either the short-eared intense female was sired by a short-eared intense male introduced into the cage or else the short-eared intense female, herself a purebred for both characters, was placed in the cage when a nurseling.

In the event that an intensely pigmented sire had been introduced into the cage by mistake, it could not have escaped and would have been observed. Furthermore, if a short-eared intense mouse were the father of the litter in which the mutant was found, all of the young would have been intense. All save the female mentioned were dilute. We might also expect other females in the cage to have been fertilized by the same foreign male, producing other litters containing irregularities. However, no other unexpected characters appeared.

Had the intensely pigmented female herself been introduced into the cage as a mouseling, the chances are that she would have differed from the rest of the litter in size and genetic characters other than intensity of pigmentation. She was first observed at about two weeks of age when any size variation due to difference in age should have been apparent. None was found. In addition, she exhibited no characters other than those common to the short-eared dilute strain. Her breeding tests afford convincing evidence that she could not have been introduced into the cage containing the short-eared dilute stock, but actually belonged to the litter in which she was found. She was homozygous for short-ears and carried as a recessive the dilution factor, showing that in at least one of a pair of allelomorphic chromosomes she bore the almost inseparably linked genes, short-ears and dilution, which combination could only have been derived from the short-eared dilute stock itself.

We may imagine several physical changes in the germ plasm which could have produced such an intensely pigmented mouse: a gene mutation causing completion of the pigment forming reaction ordinarily interrupted by the presence of the dilute gene in homozygous form or, perhaps, unequal crossing-over. However, the most probable explanation seems to be that the unexpected appearance of intense pigmentation was due to a reverse mutation which had occurred either somatically in the female herself or germinally in one of her immediate parents.

The mutant proved to have a very delicate constitution and to be an extremely poor breeder. Throughout the greater part of her life she was emaciated and suffered severely from pneumonia. At weaning she was mated to a dilute male and at the age of six months she produced a litter of three undernourished young. At birth these were fostered to a healthy nurse. One died shortly, but the other two lived to develop dilute pigmentation.

This result does not provide the complete information we desire, but merely tells us that the mutant was capable of transmitting the dilute gene. We should like to know, however, whether all her gametes carried the dilute gene or whether half of them bore the intense. If the former alternative obtained, the mutation was somatic. If the latter was true, then the mutation took place in the line of germ cell divisions.

For more than a month subsequent to the birth of her dilute young, the mutant was so feeble that all hope was given up for her recovery. After this, however, she rallied sufficiently to produce a second litter of three. These all developed intense pigmentation, proving the mutation to have been germinal and to have involved probably one only of the two gametes from which she was derived.

TRISOMIC INHERITANCE OF DOUBLENESS, COMPLICATED BY LETHALS, IN MATTHIOLA INCANA[1]

BY HOWARD B. FROST

UNIVERSITY OF CALIFORNIA, RIVERSIDE, CALIF.

Communicated June 16, 1931

It is a matter of peculiar interest to find evidence of trisomic inheritance of doubleness in *Matthiola*, a character whose exceptional behavior in disomic inheritance was held by Bateson to be evidence against the general occurrence of genetic segregation at meiosis.[2] Doubleness is recessive and doubles are completely sterile. In a "double-thrower" (or "ever-sporting") race, as Miss Saunders has shown,[3] all singles are heterozygous for double; their functional sperms transmit double only, and their eggs, both single and double, but their double progeny from selfing slightly exceed the 50 per cent expected from a simple back cross; in F_2 from crosses of double-thrower with pure single, diploid F_1 plants which carry doubleness usually give about 25 per cent of doubles (that is, these F_1's are ordinary heterozygotes, not double-throwers).

The genic formulation for doubleness developed by Haldane and Waddington[4] will be mainly employed, although the dihybrid ($XYxy$) scheme of Miss Saunders,[3] as completed by Muller with two pollen lethals,[5] gives similar expectations in large part. In view of the fact that the symbols Ll had previously been used for other genes,[6] Waddington suggests in correspondence the substitution of l_s and l_w for his published v and l. On this basis the normal or "wild-type" constitution is $\frac{L_sSL_wW}{L_sSL_wW}$ (homozygous for single flowers, colorless petal plastids, and absence of lethals); and the ordinary diploid double-thrower single of a sulfur-white race is $\frac{l_sSl_wW}{L_ssL_ww}$, the recessive mutant genes being: l_s, pollen lethal,[4] and also egg lethal unless l_w is present; l_w, pollen lethal; s, double (sterile) flowers; w, cream plastids. The dominants will hereafter be omitted in

the zygotic formulas; those just given may then be written $+/+$ and $l_s l_w / s_w$. Crossing-over is assumed to occur within this group only in the $s - l_w$ region. As the data to be reported here do not involve cream flower color, the genes *Ww* will be omitted in the following discussion.

The Crenate trisomic occurs frequently among the progeny of normal (diploid) parents of the long-chromosome variety "Snowflake," while Slender seems much rarer.[6,7]

Crenate ($7_{II} + Cr$) has a large extra chromosome, presumably complete;

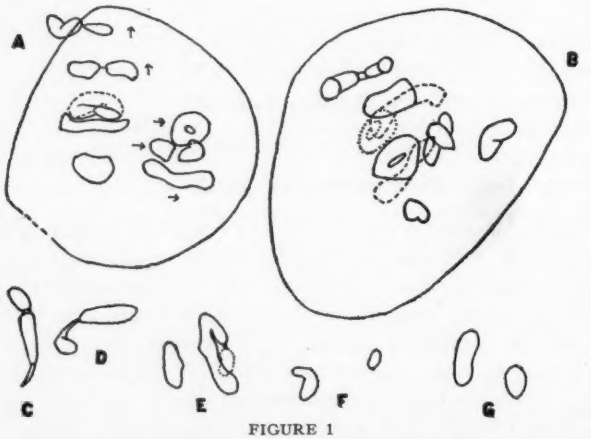


FIGURE 1

Chromosomes in pollen mother cells of one hybrid *CrSl* plant; iron-acetocarmine; camera lucida, Zeiss 70X water-immersion objective, 25X ocular; about 3050 diameters, reduced one-third in reproduction. All at first metaphase, except *C* and *D* slightly earlier. *A:* $7_{II} + 1_{II}$ (*Cr* and *Sl*, apparently joined), chromosomes on two levels displaced in drawing as indicated. *B:* $7_{II} + 1_I$ *Cr* $+ 1_I$ *Sl*. *C–G:* behavior of the extra chromosomes of five pollen mother cells; *C,D:* 1_{II} *CrSl;* E: 1_{III} *CrCrSl* $+ 1_I$ *Cr;* *F,G:* 1_I*Cr* $+ 1_I$ *Sl;* in *E, F* and *G*, the separate chromosomes shown relatively about half as far apart, horizontally, as in the original drawings.

trivalents frequently occur at the first metaphase, although, as in the other *Matthiola* trisomics examined, the extra chromosome is unconjugated in a very large part of the pollen mother cells. Slender ($7_{II} + Sl$) has a much smaller extra chromosome, evidently a fragment, and trivalents seem to be much less common than with Crenate.[8] An unconjugated extra chromosome in *Matthiola* may either divide in the first division and assort in the second, or vice versa.[9]

In the plants of the incompletely tetrasomic form, Crenate Slender ($7_{II} + Cr + Sl$), the two extra chromosomes are readily distinguished

by their relative size (Fig. 1); that they are homologous is indicated by the fact that in one of the two CrSl plants examined they often constituted an eighth pair at the first metaphase in the pollen mother cells. Of 52 counted groups from this plant, 18 showed only eight chromosomes, all clearly or probably bivalent; in 5 of these 18, one bivalent was visibly composed of two elements very unequal in size (Fig. 1, A, C, D).

A first metaphase of Slender, in which both whole *Cr* chromosomes go to the same pole and the *Sl* chromosome to the other pole, would give one Crenate-producing second-metaphase group ($7_I + Cr$), while the other second-metaphase group ($6_I + Sl$) would presumably produce non-functional pollen. The fact that Crenate mutants are only occasional among the progeny of Slender parents, indicates that in Slender the two whole chromosomes of the trisome usually disjoin (segregate) at meiosis.

Therefore, if the extra chromosome of Crenate carries the *Ss* locus, it may be anticipated that Crenate parents will give doubleness ratios corresponding more or less closely to random segregation in the trisome, but that with Slender the ratios will diverge markedly from the random trisomic expectation.

The selfing ratios given by Crenate and Slender double-throwers indicate that these singles are always *Sss*, never *SSs*.[7,10] Since there is evidence[10] that within a double-thrower race functional pollen may carry an *S* chromosome, provided an *s* chromosome is also present, it may be assumed that at least the constitution $s/l_s l_w/l_s l_w$ is zygotically lethal. It will be tentatively assumed that any trisomic constitution including either $L_s l_s l_s$ or $L_w l_w l_w$ is zygotically lethal, although the latter combination may be viable.

When double-thrower Crenate ($2n + Cr$) or Slender ($2n + Sl$) is crossed with a pure single (*SS*) normal ($2n$), therefore, all F_1 plants receiving the extra chromosome are expected to be either *Sss* or *SSs*. If we assume that the Crenate double-thrower parent had the constitution $s/s/l_s l_w$, the most probable *Sss* constitution among the F_1 hybrids will be $+/s/s$; the expected F_2 ratios will then be: normals, 5 single: 4 double; trisomics 7 single: 2 double. Three F_1 Crenate parents (table 1, parents Cr 1 to Cr 3) gave ratios corresponding well with this expectation.

The calculations for both Crenate and Slender assume, primarily, equal transmission of the extra chromosome by eggs and sperms. Actually the male transmission appears to be only about one-fifth of the female in Crenate, and one-half in Slender. In cases where unequal transmission would modify the ratio, therefore, the expectation for transmission through the eggs alone is added in parenthesis.

On the assumptions stated above, the only possible crossover *Sss* constitution in F_1 is $+/s/sl_w$. The F_2 expectation here, assuming that crossing-over between $L_s S L_w$ and $L_s s l_w$ chromosomes cannot occur,[4]

and ignoring the possibility of differential viability, is: normals, 2 single : 1 double; trisomics, 11 single : 2 double (for egg transmission, 5 : 1). The F_2 Crenate parent Cr 4 (table 1) gave practically 2 : 1 among the normals; among the trisomics there was a small excess of doubles.

TABLE 1

SINGLE AND DOUBLE PROGENY OF CRENATE TRISOMICS ($2n + Cr$) AND OF RELATED NORMALS ($2n$); PARENTS SNOWFLAKE (DOUBLE-THROWER) OR F_1 HYBRIDS BETWEEN SNOWFLAKE AND PURE SINGLE RACES (NUMBERS IN PARENTHESIS CALCULATED FROM PROBABLE CONSTITUTION)

RACE AND TRISOMIC TYPE	PARENTAGE PROBABLE CONSTITUTION	INDIVIDUALS TESTED*	NORMAL PROGENY SINGLE	NORMAL PROGENY DOUBLE	TRISOMIC CRENATE PROGENY SINGLE	TRISOMIC CRENATE PROGENY DOUBLE
Snowflake, Crenate, selfed	$s/s/l_s l_w$	Many	92	413	150	57
Snowflake, Crenate	Crenate X					
X normal (diploid)	$s/l_s l_w$	Several	4	37	5	4
Snowflake, normal	$s/l_s l_w X$					
X Crenate	Crenate	Several	143	172	10	7
	$+/s/s$	Cr 1	69	58	41	10
	Same	Cr 2	46	47	36	8
	Same	Cr 3	105	78	43	16
Hybrid, F_1, Crenate, selfed	Same	Cr 1 to	220	183	120	34
		Cr 3	(224)	(179)	(120)	(34)
	$+/s/sl_w$	Cr 4	99	52	82	24
			(101)	(50)	(90)**	(16)**
	$+/s/l_s l_w$	Cr 5	147	33	42	1
			(150)	(30)	(43)	(0)
	Same ?	Cr 6	45	4	10	0
		N 1	69	15	0	0
		N 2	59	18	0	0
		N 3	108	31	0	0
Hybrid, F_1, normal, selfed	$+/s$	N 4	104	26	1	0
		N 5	93	31	0	0
		N 6	126	43	0	0
		N 1 to	559	164	1	0
		N 6	(542)	(181)	(0)	(0)

* "*Cr*" indicates Crenate trisomics; "N" indicates normals. Parents Cr 1, N 1 and N 2 came from one varietal cross, Cr 2, Cr 4, N 3 and N 4 from a second and the other hybrid parents from a third; except that Cr 6 came from a fourth cross. All hybrid Crenate parents except Cr 2 and Cr 4 were new chromosome mutants in F_1, but probably received the extra chromosome from the Snowflake (long-chromosome) parent.[6]

** With egg transmission only, 88 : 18.

From the parentage assumed, the most probable *SSs* constitution in F_1 is $+/s/l_s l_w$, corresponding to an expectation (without crossing-over) of: normals, 5 single : 1 double; trisomics, all single. Crossing-over between $L_s L_w$ and $l_s S l_w$ would slightly increase the proportion of normal singles. The progeny of Cr 5 approximate 5 : 1 and 1 : 0 very closely;

the one exceptional trisomic can be explained either by crossing-over in the four-strand stage or by double non-disjunction in the second division without crossing-over. Parent Cr 6 was inadequately tested, from poor seed; its constitution may have been $+/s/l_s l_w$, or possibly $+/sl_w/l_s$.

The six comparable normal F_1 hybrids tested have given ordinary disomic ratios. Altogether, the results agree well with expectation on the assumption that the *Ss* locus was present in triplicate in the hybrid Crenates, with random reduction in the trisome.

For double-thrower Crenate, which might be either $s/s/l_s l_w$, $s/sl_w/l_s$ or $s/s/l_s$, the available data are summarized in table 1. Without crossing-over, the ratios expected from selfing of $s/s/l_s l_w$ are: normals, 1 single:2 double; trisomics, 2.25 single:1 double (egg transmission, 2:1). If the $l_s Sl_w$ chromosome crosses over in the s-l_w region with each $L_s sL_w$ chromosome in 6 per cent of the sporocytes (about the rate indicated for disomic crossing-over by Miss Saunders' results), the ratios become 1:2.27 and 2.15:1 (2:1). A lower rate of crossing-over than in the one pairing of the disomic would be expected. Selfing of $s/sl_w/l_s$ similarly gives 0:1 and 2.5:1 (2:1), changed by crossing-over to 1:33.33 and 2.36:1 (2:1). And $s/s/l_s$ (lacking the lethal l_w) would give only 0:1 and 2:1, single normals being entirely absent.

Obviously, both here and with the hybrid Crenates, there is much elimination of the extra chromosome, which (table 1) occurs very largely in the pollen mother cells or in the pollen. Extended consideration of these ratios has suggested no plausible complete explanation (aside from the improbable one of sampling error) that does not also involve both non-random segregation (widening both ratios) and selective elimination of singleness among the Crenates—the two complications suggested by Waddington for Slender. Probably the two similar (*s*) chromosomes conjugate more readily with each other than with the lethal-bearing *S* chromosome. The selective elimination of single may not be entirely confined to the Crenates, or to the haploid stage, however; there is evidence suggesting a slightly lower viability of single, probably in part zygotic, in ordinary double-thrower races.[3a,11,12] Such an effect may well be much more marked in the weak and relatively infertile Crenate parents, and among the Crenate haplonts and embryos.

Waddington's hypothesis will be applied to the ratios from selfing of Crenate double-thrower, assuming the constitution $s/s/l_s l_w$. If the ratio of the pairing between the two *s* chromosomes, to that between the *S* and *s* chromosomes (assuming in each case, for simplicity, that the third chromosome assorts independently), is represented by m:2 (in random assortment, $m = 1$), and if p gametes of the constitution $L_s sL_w/l_s Sl_w$ survive for every one $L_s sL_w/L_s sL_w$ gamete, the progeny ratios of singles to doubles (assuming equal transmission and no crossing-over) will be:

diploid, $1:m+1$; trisomic, $p(m+1)+\frac{1}{2(m+1)}:1$. The two observed doubleness ratios are precisely fitted if we take $m = 3.49$ and $p = 0.56$. With egg transmission only, the expected trisomic ratio is $p(m+1):1$ and $p = 0.59$.

The data (mainly published[7]) might represent the constitution $s/s/l_s l_w$, and perhaps occasionally $s/sl_w/l_s$.

The two constitutions that are possible for double-thrower Crenate

TABLE 2

SINGLE AND DOUBLE PROGENY OF SLENDER TRISOMICS ($2n + Sl$); PARENTS SNOWFLAKE (DOUBLE-THROWER) OR HYBRIDS BETWEEN SNOWFLAKE AND PURE SINGLE RACES, ALL SELFED EXCEPT AS INDICATED (THESE ARE SELECTED TYPICAL CASES)

RACE AND TRISOMIC TYPE	PARENTAGE POSSIBLE CONSTITUTION OF SLENDER	INDIVIDUAL TESTED*	NORMAL PROGENY SINGLE	NORMAL PROGENY DOUBLE	TRISOMIC SLENDER PROGENY SINGLE	TRISOMIC SLENDER PROGENY DOUBLE
	$s/s/(l_s l_w)$	Sl 1	6	98	34	8
	Same	Sl 2	1	53	26	2
Snowflake,	Same	Sl 2 X Crenate**	5	65	26	3
Slender	(X s pollen)					
	Same	Sl 3	5	122	36	12
	Same	Sl 1 to Sl 3, selfed†	12	273	96	22
	$s/l_s l_w/(s)$	Sl 4 to Sl 7†	35	94	80	27
	$+/s/(s)$	Sl 8	73	20	43	16
Hybrid, F_1,	Same	Sl 9	70	24	51	17
Slender	$+/s/(l_s l_w)$	Sl 10	75	30	37	0
	$+/sl_w/(s)$ or $+/l_s l_w/(s)$	Sl 11	136	2	132	1

* "Sl" indicates Slender trisomics. Hybrid parents Sl 8 and Sl 10 came from the same varietal cross as Cr 3 and Cr 5 of table 1; parents Sl 9 and Sl 11, from the same varietal cross as Cr 2 and Cr 4 of table 1. The pollen parent of Sl 8 and Sl 10, and the seed parent of Sl 9, was Sl 2; the pollen parent of Sl 11 was another Snowflake Slender.

** The functional pollen of Crenate, except for a small percentage producing Crenate progeny, must be $L_s SL_w$. The expectation is very nearly the same as with selfing for this constitution of Slender (see text).

† The observed percentages of double Slender cited by Waddington[4] are exaggerated by an arithmetical slip in his calculations. The value for selfing should be 28.6 (not 40) per cent.

on the $XYxy$ scheme, under the corresponding assumptions as to lethals, give (with crossing-over) either similar or less promising expected ratios, as follows: (1) normals, 1:2.41, and trisomics, 1.78:1 (egg transmission, 1.68:1); (2) 1:49 and 1:1.05 (1:1.83).

Slender is much more vigorous and fertile than Crenate, and consequently the genetic tests have been much more extensive; selected typical results are reported in table 2. The main features of the cytology and transmission of the Slender type, as previously described, preclude the

application of the hypothesis of selective segregation, *in the simple form stated by Waddington*, to these ratios.

That parent Sl 1 of table 2 had the constitution $ss(S)$ (that is, with the one S carried in the Sl fragment) is shown by the fact that it produced a variant branch with some fully double flowers and normal-appearing leaves; loss of the Sl chromosome obviously removed the only S gene carried by this plant.[13] Cytological examination of several similar cases of somatic "mutation" (not involving doubleness) has confirmed this interpretation of the nature of the change. The ratios indicate that all other Snowflake Slenders adequately tested (for example, Sl 2 and Sl 3 of table 2) carried only one S.

If the chromosomes have their usual diploid constitution, and there is no crossing-over involving the fragment, an $ss(S)$ double-thrower Slender parent should produce only double normals and single Slenders. On the same conditions, and in the absence of selective segregation or elimination, an $sS(s)$ Slender should give the ordinary ratio of 1 single:1+ double among the normal progeny, and 3 single:2+ double (by egg transmission, 1:1+) among the trisomic progeny. Neither of these results has been secured, but most of the observed ratios suggest the former constitution with considerable crossing-over (table 2). This constitution may be represented by the formula $s/s/(l_s l_w)$. As will be shown below, the expectation is nearly the same if l_w changes places with an L_w in this constitution.

With the constitution $s/s/(l_s l_w)$, if the deficiency of the Sl-chromosome lies to the right of the l_w locus, crossing-over in the s-l_w region, involving the fragment, would produce Slender doubles, but normal singles would be absent because of the egg-lethal effect of $l_s L_w$. Crossing-over to the right of l_w (that is, beyond the completely linked W_w locus) would give both cross-over classes. With the constitution $s/sl_w/(l_s)$, on the other hand, normal singles could be produced only by crossing-over in the former region.

The ratios given by the selfing totals for parents Sl 1 to Sl 3 could be produced, with the first constitution, by crossing-over alone (14 per cent in the former region and 4 per cent in the latter, in each pairing, giving: normals, 13 single:272 double; trisomics, 97:21). So high a rate in the former region is improbable, however, and it seems more likely that selective elimination of trisomic single has a marked effect on this ratio. The elimination of the extra chromosome has been somewhat similar to that with Crenate, although considerably less severe in the pollen.[10]

The assumption of 6 per cent of crossing-over in each pairing in the s-l_w region, with 4 per cent in the other region, gives the ratios: normals, 1:22.5; trisomics, 9.01:1 (egg transmission, 9:1), the ratio among the trisomics requiring extensive selective elimination of trisomic single in

order to correspond to the observed ratios. Unless otherwise stated, these rates of crossing-over are to be understood whenever crossing-over is mentioned below.

Most of the progeny tests agree with the sample cases reported in table 2 in suggesting the parental constitution $ss(S)$. Such ratios are fitted about equally well whether the lethal l_w is in the fragment or in a whole chromosome; the constitution $s/sl_w/(l_s)$ gives, with crossing-over, doubleness ratios (normals, 1:30; trisomics, 9.79:1; with egg transmission only, trisomics, 9:1) not readily to be distinguished from those given by the first constitution above. One $sS(s)$ constitution, $sl_w/l_s/(s)$, gives somewhat similar ratios, 1:18.33 and 3.03:1 (egg transmission, 1.22:1), with crossing-over; without crossing-over, normals would be entirely absent (and with egg transmission only, trisomics also absent).

If the deficiency of the *Sl* fragment is to the *left* of l_s, the ratios with crossing-over are: normals, 1:24; trisomics, 23.30:1 (24:1), for the first constitution discussed, 0:1 and 28.96:1 (24:1) for the second, and nearly as above for the third (egg transmission, 1.08:1). Therefore it seems more probable that the missing region is located to the right of l_w (and hence to the right of w), and that location is generally assumed in this discussion. This inference agrees with the assumption that l_s is the deficiency which Philp and Huskins have demonstrated cytologically.[4]

The disagreement of the earlier data for selfed Slender double-thrower[10] with those just considered is due mainly to small-scale data from four parents (Sl 4 to Sl 7), which gave closely accordant results totaling: normals, 35:94; trisomic Slenders, 80:27; the remaining parents gave 7:195 and 60:29. This suggests for the former parents the constitution $s/l_sl_w/(s)$, which without crossing-over would give 1:1 and 1.5:1 (1:1), and with crossing-over would give 1:1.31 and 1.54:1 (1.22:1). No other constitution that is possible under the present assumptions gives an expectation even remotely approaching the observed narrow ratio for normals that is here in question. *This* constitution, however, might give such a ratio as a result of selective segregation such as Waddington has suggested; if (L_sL_w) separates from L_sL_w much more often than from l_sSl_w, the ratios would be much widened. The expected ratio for the trisomics, however, does not leave room for the marked selective elimination of single which would be anticipated.

Of the seven constitutions permitted to double-thrower Slender by our hypothesis, another that might be expected to be common is $s/l_s/(sl_w)$; without crossing-over this gives: normals, 0:1; trisomics, 1:1; and with crossing-over: 0:1 and 1.23:1 (1.22:1)—ratios not suggested by any results. There are two possible constitutions carrying only one lethal (l_s); the ratios with crossing-over are 0:1 and 9:1 from $s/s/(l_s)$ and 0:1 and 1.22:1 from $s/l_s/(s)$.

Under the limitations imposed above, the *XYxy* scheme permits five constitutions of double-thrower Slender parents, corresponding respectively to the first five *Ss* constitutions above; the expectations with crossing-over are very similar in the first, fourth and fifth cases, and somewhat similar in the third (1:32.33 and 1.94:1; egg transmission, trisomics 1:9); in the second case, the ratio among the trisomics is only 1.30:1 (1.08:1). On account of the second case, the *XYxy* scheme seems to correspond to the results somewhat less well than the *Ss* scheme.

Most if not all of the F_1 hybrid Slenders tested seem to fall into three classes. Often the F_2 progeny ratios approach 3 single:1double for both normal and trisomic progeny (table 2, parents Sl 8 and Sl 9). This suggests the F_1 constitution $+/s/(s)$, which indicates an $s/(s)$ gamete from the P_1 Snowflake Slender. Another F_2 result (Sl 10) approaches: normals, 3:1; trisomics, 1:0. This suggests $+/s/(l_s l_w)$, the constitution which we should expect to find much the most often if the P_1 Slender was $s/s/(l_s l_w)$. The third result (*Sl* 11) approaches entire absence of doubles; this is common, and has been obtained both when the double-thrower Slender was seed parent and when it was pollen parent in the cross. This suggests the presence of one or both lethals in the whole chromosome derived from the P_1 Slender, as is indicated in table 2.

The suggestion is obvious that Miss Saunders' cases of marked deficiency of doubles, in F_2 from crosses with pure single,[3a] may be due to a lethal in the *s* chromosome in F_1—apparently a lethal other than l_w, since $L_s SL_w/L_s sl_w$ ($= XY/Xy$) seems to give no recombination for plastid color.[3b] Evidence for this interpretation is furnished by one case in my cultures. Of two F_1 sibs, belonging to a trisomic type (Smooth) that is disomic for the Cr (doubleness) chromosomes, one gave in F_2 about 3 single:1 double, while the other gave only about 3 per cent of doubles. A gene for recessive intense sap color (p)[14] had entered the cross in the *s* chromosome, and the deficiency of double was accompanied by a similar deficiency of intense.

A thorough analysis of the genetic behavior of doubleness in these trisomic forms, which would necessarily include a study of the linked gene pairs *Ww* and *Pp*, would obviously require many times as many plants as the few thousands grown in the work here discussed. The correspondence of result with theory has been sufficiently extensive and varied, however, to encourage the hope that the main features of an adequate explanation have been indicated by the evidence at hand.

A point of especial theoretical interest is the clear indication, from crosses in which Slender was pollen parent,[10] that in a double-thrower race functional pollen can carry a singleness chromosome, or at least a portion of it which includes the singleness locus, *provided* it also carries a doubleness chromosome. This fact tends to confirm the gamete–lethal

hypothesis for the explanation of the absence of singleness from ordinary double-thrower pollen, as also does the indicated non-occurrence of *SSs* double-thrower Slenders in spite of the abundance of *S* eggs and *Ss* sperms.

The frequent occurrence of crossing-over involving a chromosome fragment (*Sl*) seems unquestionable.

[1] Paper No. 253, University of California, Graduate School of Tropical Agriculture and Citrus Experiment Station, Riverside, California.

[2] Bateson, W., *Science*, n. s., **40**, 287–302, 1914.

[3] Saunders, E. R., (*a*) *J. Genetics*, **1**, 303–376, 1911; (*b*) **20**, 53–77, 1928; (*c*) *Am. Nat.*, **50**, 486–498, 1916.

[4] Waddington, C. H., *J. Genetics*, **21**, 193–206, 1929. On the basis of Snow's remark (*Am. Nat.*, **58**, 316–321, 1924) about Haldane's hypothesis, I independently (although later) applied to Miss Saunders' results the same scheme as did Waddington, except for a more inclusive tentative hypothesis for the apparent limitation of crossing-over between the *Ss* and L_wl_w loci. I assume that the lethal l_w (or some condition associated with its presence) inhibits crossing-over of its chromosome with any L_sSL_w ("wild-type") chromosome, but not with L_sSL_w. This assumption covers all the cases (matings 1, 2 and 7 of Miss Saunders' tabular analysis)[3b] where there is some reason for supposing that crossing-over does not occur, while Waddington's assumption covers only mating 7. Miss Saunders' data for the 1:1 ratio from mating 1 (783 white : 879 cream; p. 55, 56) give odds of about 3845:1 against sampling deviation from Waddington's 16:15 ratio, but only 52:1 against the 1:1 ratio expected without crossing-over; however, selective elimination of the lethal-bearing white class might account for either deviation. The work of Philp and Huskins indicates that the lethal l_s consists of the absence of a satellite normally present (Philp, J., and C. L. Huskins, *Anat. Record*, **47**, 379, 1930, Abstract).

[5] Muller, H. J., *Genetics*, **3**, 422–499, 1918.

[6] Lesley, M. M., and H. B. Frost, *Genetics*, **12**, 449–460, 1927.

[7] Frost, H. B., *Univ. Calif. Publ. Agric. Sci.*, **2**, 81–190, 1919.

[8] Frost, H. B., and M. C. Mann, *Am. Nat.*, **58**, 569–572, 1924. Frost, H. B., *J. Heredity*, **18**, 474–486, 1927. The cytological evidence mentioned in the present paper was mainly obtained by Dr. Margaret Mann Lesley, although the author obtained the evidence here presented relating to the behavior of the two extra chromosomes in a Crenate Slender plant. All general conclusions as to amount of conjunction of extra chromosomes in *Matthiola* are tentative, since as a rule relatively few pollen mother cells permit reliable counts, and doubtless those in which the extra chromosome is separate are more often countable. The chromosomes of the *Ss* pair may in general be designated *A* (instead of *I* as previously;[7] for simplicity, the trisomic symbol *Cr* has been used throughout the present discussion. Trisomic Crenate might be called Triplo-A.

[9] Lesley, M. M., and H. B. Frost, *Am. Nat.*, **62**, 22–33, 1927.

[10] Frost, H. B., *J. Heredity*, **19**, 104–111, 1928.

[11] Saunders, E. R., *J. Genetics*, **5**, 137–143, 1915.

[12] Frost, H. B., *Am. Nat.*, **49**, 623–636, 1915. Miss Saunders' totals of 1911 for selfing and inter-crossing of double-throwers,[3a] omitting families indicated as from old seed, 13,631 plants altogether, show, according to my calculations, 56.39 ± 0.29 per cent of doubles, while the percentage expected on her genic scheme, with 6 per cent of crossing-over, is only 53 per cent, and on Waddington's scheme only 51.5 per cent. Snowflake has given 52.90 ± 0.42 per cent.[10] It should be noted that Miss Saunders' scheme requires crossing-over between the *Ww* and single-double loci in the *X*-*Y* region

alone in double-throwers, and in the *W-X* region alone (when at all) in hybrids with pure single, but the indicated rates for the two regions are identical or similar. Renner's evidence relating to megaspore competition in *Oenothera* indicates that megaspore development should be studied in *Matthiola* (Sturtevant, A. H., *Quart. Rev. Biol.*, **1**, 283–288, 1926.

[13] Frost, H. B., *J. Agric. Research*, **33**, 41–46, 1926.

[14] Saunders, E. R., *4e Conf. int. Génétique, Compt. Rend.*, 397–405, 1913.